PRECISE GENOME EDITING HUMAN STEM CELLS FOR FUNCTIONAL INTERROGATION OF GENETIC VARIATION

By
Cory James Smith

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore Maryland

June 2014

ABSTRACT

Human induced pluripotent stem cells (iPSCs) provide great potential as a tool for basic biological discovery, disease modeling, and ultimately for cell based regenerative medicine but the full realization of these stem cells requires an ability to precisely edit their genome in an efficient way. The relative efficiencies of CRISPR/Cas9 and TALENs were quantified in human iPSC lines for inducing both homologous donor-based precise genome editing and nonhomologous end joining (NHEJ)-mediated gene disruption; genome wide off-target mutagenesis was also assessed by targeted deep sequencing. The specificity of Cas9 was further tested by targeting either the mutant or the wild-type allele of disease causing single nucleotide variants (SNVs) with gRNAs and testing their cleavage at the intended target or the other genotype differing by a single nucleotide; little disruption was observed at the other allele differing by a single nucleotide alone. Overall, these results demonstrate the advantages of the CRISPR/Cas9 system in allele-specific genome targeting and in NHEJ-mediated gene disruption while they were comparable in HDR efficiencies. To further investigate the specificity whole genome sequencing was performed on iPSCs successfully edited with either CRISPR/Cas9 or TALENs and found no evidence of mutations at other sites similar to the nuclease binding sequence. To utilize these genome editing tools for a relevant biological phenotype, genetic variation associated to platelet aggregation was chosen for further characterization. The platelet endothelial aggregation receptor 1 (PEAR1) gene was knocked out (KO) using CRISPR/Cas9 and compared to the otherwise isogenic iPSCs to determine its influence on hematopoiesis and megakaryopoiesis. PEAR1 KO iPSCs were

found to increase cell proliferation as well as increase megakaryocyte lineage commitment. Additionally an intronic SNP, rs12041331, reported to reside within a putative enhancer was deleted along with the surrounding 251 bp which resulted in a significant reduction of PEAR1 mRNA levels. This reduction was primarily observed on the same allele as the deletion indicating a cis-acting mechanism of gene regulation. Overall this thesis demonstrates the efficiency and specificity of genome editing in human stem cells and their use towards precise genetic modification for functional interrogation of genetic variation.

Advisor: Linzhao Cheng
Reader: Gabsang Lee

Preface

Acknowledgements

I would first foremost like to thank my mentor Dr. Linzhao Cheng for providing me with the resources, lab environment, and training for me to make the transition from a student into a laboratory scientist. I will always be appreciative of my time spent here and the transformation that occurred both for my career and personally. I would like to thank key members of the Cheng lab who taught me the techniques of genome editing in human pluripotent stem cells including Dr. Prashant Mali, Dr. Jizhong Zou, and Dr. Zhaohui Ye. I would also like to thank Dr. Kun Zhang and Dr. Athurva Gore in UCSD for their help in library preparation, and whole genome sequencing and analysis conduced in aim two to investigate genomic integrity after editing. Additionally I would like to thank Dr. Hans Bjornson and Dr. Li Zhang for training and allowing me to use their MiSeq for targeted deep sequencing used throughout the thesis. I would also like to thank my thesis committee members including Dr. Steven Salzberg, Dr. Kathy Burns, Dr. Gabsang Lee, and Dr. Zhaohui Ye for the continued mentorship and advice throughout my research here and into the future.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

During the natural course of human embryogenesis the single cell composing the zygote contains all the genetic information and potential to produce an adult human capable of perpetuating the cycle once more. After a certain stage of embryogenesis no pluripotent stem cells remain with the potential to differentiate into all cells of the body leading to a state of minimal self-repair that cannot recover from severe injury or disease. While the sequence of DNA in the nucleus has not changed in the adult, its epigenetic landscape has been corralled into a Waddington final state incapable of return to the summit, or so it seemed until cellular reprogramming was discovered first by John Gurdon in frogs. In 1958 Gurdon first successfully cloned a Xenopus by transferring a nucleus derived from a tadpole into an enucleated egg (Gurdon et al. 1958). Then some set of unknown factors in the nucleus of the egg "reprogrammed" the nucleus of the somatic cell back to an embryonic state that was capable of populating the entire frog. For decades the work was not reproduced by others and largely ignored until Dolly the sheep was cloned by Somatic Cell Nuclear Transfer (SCNT) proving this technique was also applicable to mammals. Using a brute force reductionist approach Shinya Yamanaka found that four simple transcription factors, OCT4, SOX2, KLF4 and cMYC could reprogram an adult cell back to an embryonic like state capable of self renewal and the capacity for differentiation into all cell types that he called induced Pluripotent Stem Cells (iPSCs) (Takahashi and Yamanaka 2006). These iPSCs are a tremendously powerful tool to contribute to basic biological research, disease modeling and have potential applications in autologous cell based regenerative medicine.

DNA sequencing technology has been steadily improved over the past four decades leading to the throughput necessary to sequence a draft version of the human genome (Venter 2001; Lander 2001). The cost of sequencing continued to drop with the development of new methods for DNA sequencing that include pyrosequencing, sequencing by ligation, and reversible terminator chemistry that all implemented massively parallel techniques to dramatically increase the throughput and cost efficiency of reading DNA. This ushered in a new era of DNA sequencing both for basic biological research as well as many new tests for genetic disorders. After more than three billions years of evolution into ever more complex chemical, cellular, and technological creatures we are apertures for the cosmos to know itself; which for the first time has gazed back upon the chemical nature of our transition from the inanimate into the living through information contained in DNA. Starstuff powered by starlight seeking to better understand where we came from to help facilitate our path into the future.

Horizontal gene transfer is the most diverse and abundant tool by which evolution has used time and time again to manipulate the genomes of all life on Earth. By mimicking simple natural systems man was able to use viruses and other gene delivery systems to insert foreign DNA into bacteria, plants and animals. Although these methods could be quite efficient at gene delivery, controlling the site and number on integrations in the host cell was problematic and could lead to transgene silencing or oncogenic transformation in the case of *in vivo* therapies. Genome editing using a homologous DNA donor relies on spontaneous generation of DNA double strand breaks (DSB) to recombine the exogenous donor template with genomic DNA to induce a desired

mutation and has been widely used in mouse and many other model organisms but has been difficult to use in human stem cells due to poor survival of single cells, and orders of magnitude lower efficiency of recombination. To increase the efficiency of recombination DNA binding domains such as Zinc Finger domains were combined with a Fok1 nuclease domain to generate Zinc Finger Nucleases (ZFNs) which allowed successful gene targeting at several endogenous loci in the human genome (Zou and Maeder et al. 2009; Zou and Mali et al. 2011). Although ZFNs greatly facilitated the genome editing process they were difficult to design and relied on companies to assembly and test many finger combinations before a final tool was developed. Transcription activator like effector nucleases (TALENs) were also developed from DNA binding domains secreted by Xanthomonas bacteria to influence plant gene expression. TALENs could target a broader range of sequences than ZFNs and were dramatically easier to synthesize and clone (Li et al. 2011; Cermak et al. 2011). The latest technology for genome editing to emerge is the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/ Cas9 derived from an adaptive bacterial immune system which uses single guide RNAs (gRNAs) to target and cleave specific DNA sequences (Jinek et al. 2012). This system has been repurposed for eukaryotic expression and has been shown effective in a broad range of animal models and human cell lines including iPSCs (Mali et al. 2013a; Cong et al. 2013).

This thesis aims to compare and contrast TALENs and CRISPR/Cas9 as a robust new tool for genome editing human stem cells as well as to perform functional interrogation of genetic variation associated with a relevant biological and medical

phenotype, platelet aggregation. The first aim investigated the specificity of Cas9 by predicting the most similar sites in the genome to the gRNA sequence and PCR amplify these sites in bulk transfected cells followed by deep sequencing to determine their indel mutation rate at sites other than the intended target; the overall efficiencies of NHEJ and HDR were also measured. The second aim probed deeper into genomic integrity after genome editing using whole genome sequencing (WGS) to evaluate the mutation rate genome wide and search for mutations similar to the DNA binding site of TALENs or Cas9/gRNA nucleases. Finally the third aim of the thesis involved applying these tools for precise genome editing towards a biological phenotype of platelet production and aggregation by modifying the platelet endothelial aggregation receptor 1 (PEAR1) gene to interrogate its function. The PEAR1 coding sequencing and putative enhancer were knocked out and compared to otherwise isogenic parental iPSCs from which they were derived to isolate specific genetic effects and determine their role on hematopoietic differentiation followed by megakaryocyte maturation. Together with next-generation DNA sequencing that uncovers genetic variations, this study demonstrated an approach and feasibility to precisely edit DNA at a selected locus and to determine the causal effects of specific DNA variants of interest.

SPECIFIC THESIS AIMS

Aim 1: Compare the efficiency and specificity of human genome editing stimulated by TALEN, and CRISPR/Cas9 in human iPSCs.

Aim 2: Investigate the genomic integrity of human iPSCs after genome editing by CRISPR/Cas9 or TALENs.

Aim 3: Knockout the PEAR1 gene and its cis-regulatory element to evaluate their function and regulation during iPSC-derived megakaryopoiesis.

Aim 1: Compare the efficiency and specificity of human genome editing stimulated by TALEN, and CRISPR/Cas9 in human iPSCs.

Efficient and precise genome editing is crucial for realizing the full research and therapeutic potential of human iPSCs. Engineered nucleases including CRISPR/Cas9 and TALENs provide powerful tools for enhancing gene-targeting efficiency. In this study, we investigated the relative efficiencies of CRISPR/Cas9 and TALENs in human iPSC lines for inducing both homologous donor-based precise genome editing and nonhomologous end joining (NHEJ)-mediated gene disruption. Significantly higher frequencies of NHEJ-mediated insertions/deletions were detected at several endogenous loci using CRISPR/ Cas9 than using TALENs, especially at nonexpressed targets in iPSCs. In contrast, comparable efficiencies of inducing homologous donor-based genome editing were observed at disease-associated loci in iPSCs. In addition, we investigated the specificity of guide RNAs used in the CRISPR/Cas9 system in targeting disease-associated point mutations in patient-specific iPSCs. Using myeloproliferative neoplasm patient-derived iPSCs that carry an acquired JAK2-V617F point mutation and α1-antitrypsin (AAT) deficiency patient-derived iPSCs that carry an inherited Z-AAT point mutation, we demonstrate that Cas9 can specifically target either the mutant or the wild-type allele with little disruption at the other allele differing by a single nucleotide. Overall, our results demonstrate the advantages of the CRISPR/Cas9 system in allele-specific genome targeting and in NHEJ-mediated gene disruption.

AIM 1 INTRODUCTION

Techniques to edit genomic DNA at a precise locus in human iPSCs present an unprecedented potential for regenerative medicine as well as disease modeling of genetic variants. The efficiency of multiple published techniques in human iPSCs has been extremely low until the development of engineered nucleases such as zinc finger nucleases (ZFNs), transcription activator like effector nucleases (TALENs), and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas systems (Zou et al. 2009; Hockemeyer et al. 2009; Hockemeyer et al. 2011; Mali et al. 2013a; Cong et al. 2013; Damian et al. 2013). ZFNs and TALENs are fusion proteins, in which Fok1 nuclease domain is fused to a DNA binding domain that can bind to and cleave a specific DNA sequence of interest. Once a targeted double strand break (DSB) has been introduced, the DNA is repaired by the cell's endogenous DNA repair machinery through one of two pathways. The error-prone NHEJ pathway often results in small insertions or deletions (indels), while the Homology-directed Repair (HDR) pathway results in precise repair with a homologous chromosome or an exogenous donor template. These engineered proteins acting as designer nucleases proved to be an efficient means to target and manipulate the genome for both gene knock out (KO) and knock in (KI) experiments. Compared to ZFNs, design of a pair of TALENs is more feasible for most laboratories and favored by many investigators. However, a pair of TALENs, each has ~15 peptide modules of 33-amino-acid units, still takes time to synthesize and test to ensure its efficiency as well as specificity.

In comparison to a pair of TALENs, the CRISPR-Cas type II is more user friendly

as the protein component (Cas9) remains the same while the short RNA components for one or multiple targets can be rapidly designed and synthesized. The system was originally identified to have three essential components: (i) Cas9 endonuclease, (ii) CRISPR RNA (crRNA) to bind the complementary DNA target, and (iii) trans-activating RNA (traRNA) to associate crRNA to Cas9. This was further reduced to two components by fusing the traRNA and crRNA to a single guide RNA (gRNA) (Jinek et al. 2012). Cas9 has been adapted for better expression in mammalian cells and was shown to be an efficient and adaptable tool in human cell lines including iPSCs (Mali et al. 2013a; Cong et al. 2013).

In order to gain a better understanding of the advantages and disadvantages of CRISPR/Cas9 and TALENs technologies in human iPSCs, we compared the efficiency of Cas9-gRNAs versus TALENs in targeting disease-associated loci. We investigated their efficiencies by measuring both NHEJ-mediated indel mutations and homologous donor-based precise gene editing. The model disease genes we used include JAK2, in which an acquired somatic point mutation (JAK2-V617F) occurs in approximately 95% of patients with polycythemia vera (PV) (Levine et al. 2008), and the SERPINA1 gene, in which an inherited point mutation (AAT Z-mutation) causes α1–antitrypsin (AAT) deficiency (Carrell et al. 2002). We also included the previously validated Cas9-gRNA and TALEN designer nucleases targeting the AAVS1 locus commonly used as a “safe harbor” in the human genome for stable transgene expression (Hockemeyer et al. 2009; Smith et al. 2008; Lombardo et al. 2011; Zou et al. 2011; Zou et al. 2012).

Concerns were raised over the specificity of the CRISPR/Cas9 system, when

reports using human cell lines detected NHEJ- mediated off-target mutations at loci with up to five mismatches to the gRNA (Hsu et al. 2013; Fu et al. 2013). As this was further investigated in other biologically relevant samples such as mouse and nonhuman primate embryos and human adult stem cells, off-target effects were not detected or minimal (Niu et al. 2014; Wu et al. 2013; Wu et al. 2014; Yang et al. 2013a; Yang et al. 2013b; Yin et al. 2014; Smith et al. 2014). In this aim we investigated the specificity of Cas9-gRNAs in human iPSCs in stimulating both NHEJ-mediated indel induction and donor-based genome editing by targeted deep sequencing of whole iPSC populations treated with Cas9-gRNAs.

The high specificity of Cas9-gRNAs also prompted us to investigate whether this technology can facilitate allele-specific targeting at point mutations in patient-specific and normal iPSCs. For this purpose, we used a variety of integration-free human iPSC lines including PV-iPSC lines that carry the JAK2-V617F mutation, AAT deficiency-iPSC lines that carry the Z-AAT mutation, and control BC1 iPSC line whose genomic integrity has been characterized in detail by next generation sequencing (Cheng et al. 2012) (Table 1). We designed gRNAs targeting either the mutant or the wild-type allele and examined their efficiency in disrupting or correcting the intended target allele or the other allele differing by a single nucleotide.

AIM 1 METHODS

Maintenance and expansion of human iPSCs

Human iPSCs were cultured with E8 medium (Life Technologies, Carlsbad, CA) on tissue culture plates coated with Matrigel (BD Biosciences, San Jose, CA) or Vitronectin (Life Technologies) as previously described (Chen et al. 2011; Wang et al. 2013). For routine passaging, iPSCs were digested with Accutase (Sigma, St. Louis, MO) for 5 minutes and washed with PBS by centrifugation at 200g for 5 minutes. Digested iPSCs were then plated at a density of 2×10^4 per cm^2 with E8 medium supplemented with 10 μM ROCK Inhibitor Y-27632 (Stemgent, Cambridge, MA) for the first 24 hours.

Expression vectors used in CRISPR/Cas9 and TALEN experiments

We used the CRIPSR/Cas9 system previously described (Mali et al. 2013a), for which an expression vector encoding humanized (h) Cas9 protein was obtained from Addgene.org (Plasmid #41815). 455-bp guide RNA (gRNA) expression cassettes including 20-bp target-specific sequence for each locus were synthesized (Integrated DNA Technology, Coralville, IA). The JAK2 and SERPINA1 gRNAs were synthesized as Gene Blocks and cloned into Zero-blunt TOPO vector (Life Technologies). The AAVS1 gRNA-T2 was obtained from Addgene.org (Plasmid #41818). TALEN constructs targeting the AAVS1 and AAT loci have been described in previous publication (Yan et al. 2013; Porteus et al. 2006). The JAK2 TALENs were constructed with the Joung Lab's REAL Assembly TALEN Kit (Addgene #1000000017) following published protocol (Reyon et al. 2012). The Cas9-gRNA and TALENs targeting each

locus were first validated using a GFP reporter system in 293T cells.

MiSeq-based deep DNA sequencing of endogenous loci after transient expression of engineered endonucleases

Human iPSCs were digested with Accutase for 5 minutes and the single cells were washed once with PBS. 2×10^6 iPSCs were then resuspended in 100 μl of P3 Primary Cell Solution (Lonza, Frederick, MD) supplemented with 2.5 μg of hCas9 plasmid and 2.5 μg of gRNA plasmid, and then nucleofected in 4D-Nucleofector (Lonza) using the hES H9 program. In TALEN experiments, 2.5 μg of each TALEN expression vector were used. The nucleofected iPSCs were then plated onto Matrigel-coated plates in E8 medium supplemented with 10 μM Y-27632. Three days after the transfection, all the cells were harvested and the genomic DNAs were isolated using DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). The genomic regions of interest were PCR amplified using Phusion DNA polymerase (New England Biolabs, Ipswich, MA). PCR products purified by PCR Purification Kit (Qiagen) were deep sequenced by MiSeq Personal Sequencer (Illumina, San Diego, CA) and demultiplexed using ea-utils program Fastq-Multx only allowing reads in which both paired ends agreed yielding an average coverage of >300,000 reads per sample. Reads for each sample were aligned using bowtie2 and the indel percentage was calculated by the number of reads with indels around 20-bp of the putative cutting site divided by the total number of aligned reads. Sequencing data was deposited at the Short Read Archive (SRA) with the accession number SRP042279.

Amplification and sequencing of putative off-target binding sites of gRNAs

To bioinformatically predict the off-target binding sites of three gRNAs (gR-AAVS1-T2, gR-JAK2-F and gR-PEAR1) a list of all sites in hg19 within five mismatches to the gRNA that were followed by a PAM sequence (NRG) was generated using the EMBOSS tool fuzznuc. The lists were ranked based on likelihood of cleavage predicted by an experimentally determined weighting algorithm (Hsu et al. 2013). Using Primer3, oligos were designed to amplify the top 15 putative off-target loci for each gRNAs which were independently transfected with Cas9 into iPSCs and 293T cells. After 96 hours, cells were harvested and gDNA was extracted with DNeasy Blood & Tissue Kit (Qiagen) and the intended targets (4) and predicted off-targets (60) were PCR amplified using Phusion DNA polymerase. Amplicons were pooled into four conditions: (i) parental iPSCs; (ii) iPSC Cas9/gRNA targeted; (iii) HEK293T parental; and (iv) HEK293T Cas9/gRNA targeted which underwent Kapa Biosystems–High-Throughput Library Prep Kit (Product # KK8234) for the MiSeq 500 cycle sequencing chemistry. Reads were aligned using bowtie2 and indel mutation rates for each of the amplicons was calculated as previously described (Hsu 2013).

Homologous donor-based gene targeting in human iPSCs

In homologous donor-based experiments targeting JAK2 or SERPINA1 loci, 2×10^6 patient-specific iPSCs cells were resuspended in 100 µl P3 Primary Cell Solution supplemented with either (i) 2.5 µg hCas9 plasmid, 2.5 µg of guide RNA plasmid and 5 µg HR donor vector; or (ii) 2.5 µg plasmid coding TALEN-left, 2.5 µg of TALEN-right and 5 µg HR donor vector. In experiments targeting AAVS1 locus, 2 million BC1 iPSCs

were nucleofected with either (i) 5 μg AAV-CAGGS-EGFP (addgene # 22212), 3 μg each of a heterodimeric TALEN pair targeting the AAVS1 (Porteus 2006) or (ii) 5 μg AAV-CAGGS-EGFP (Addgene # 22212), 3 μg hCas9 (Addgene #41815), and 3 μg AAVS1-T2 gRNA (Addgene #41815). The nucleofected iPSCs were plated onto Matrigel-coated six-well plates (nine wells per sample in AAT targeting experiments; 12 wells per sample in JAK2 and AAVS1 targeting experiments) immediately following nucleofection. 10 μmol/l ROCK Inhibitor Y-027632 was added in the E8 medium for the first 24 hour. There is no cell passaging between the initial plating and colony picking. Medium was replaced on a daily basis. Starting at 96 hours after nucleofection, puromycin (0.5 μg/ml) was added to the medium for the selection of targeted events. After puromycin selection, colonies were manually picked from different wells (e.g., two colonies were picked from each individual well in the AAT experiments) and expanded. We also tried to avoid picking colonies that grew close to each other. Genomic DNAs were isolated from the expanded clones and analyzed for targeted integration.

Statistical analyses

MiSeq sequencing data analysis was performed as previously described (Hsu et al. 2013). Statistical significance for targeted DNA amplicon sequencing was determined using the upper tail test of population proportion comparing each targeted sample to the untransfected control ($\alpha = 0.05$). To adjust for multiple comparisons the bonferroni correction was used ($P = 0.05$, $n = 29$) yielding an experiment-wide significance level of 1.72×10^{-3}.

AIM 1 RESULTS

Cas9-gRNAs are superior to TALENs in inducing NHEJ-mediated indel mutations

We evaluated the efficiency of Cas9-gRNA versus TALEN in targeting multiple endogenous loci in human iPSCs. We first determined, by MiSeq deep sequencing, the frequencies of small indels resulting from NHEJ-mediated DNA repair of DSBs caused by either Cas9 or TALENs in the absence of an exogenous donor template (Figure 1). Cas9 and the guide RNA (gR-JAK2-F) designed to target the JAK2-V617F mutation were compared to a pair of JAK2-TALENs designed to target the same JAK2-V617F region. Cas9 and the guide RNA (gR-AAT-Z) designed to target the Z-AAT mutation were compared to a pair of AAT-TALENs that previously have been shown to efficiently target the AAT Z-mutation region (Choi et al. 2013) (Figure 1-1).

Table 1 Human iPSC lines used in this study

iPSC line	Donor	JAK2-V617F allele[a]	JAK2-WT allele
BC1	Normal	0	2
PVB1.11	PV patient 1	0	2
iPV183	PV patient 2	1	1
PVB1.4	PV patient 1	2	0
PVB2.7	PV patient 3	2	0
		AAT-Z allele[b]	**AAT-M allele**
BC1	Normal	0	2
iAAT5	Z allele carrier	1	1
iAAT3	AAT deficiency	2	0

[a]Healthy control and PV-specific iPSCs with JAK2-V617F mutation. [b]Healthy control and AAT-deficiency iPSCs with AAT Z mutation.

In addition to these designer nucleases targeting disease-associated mutations, previously reported Cas9-gRNA (gR-AAVS1-T2) and TALENs targeting the AAVS1 safe harbor locus were also used in this study (Mali et al. 2013a; Porteus et al. 2006). These Cas9-gRNA or TALEN expression vectors were first tested in 293T cells using a previously reported GFP reporter system to validate their functionality. They were then transfected into three human iPSC lines with appropriate target sequences (Table 1-1). Three days after transfection, the genomic DNA was amplified by high- fidelity PCR using primers flanking the common targeted DNA region shared by both designer nucleases for each locus. MiSeq analysis of these PCR-amplified regions revealed significantly higher indel rates at all three endogenous loci after Cas9 targeting than TALENs (Figure 1-2). At all three loci, Cas9-gRNAs induced between 10- and 100-fold more indels than did TALENs in human iPSCs, reaching the level of 0.7–2.5% mutation rates.

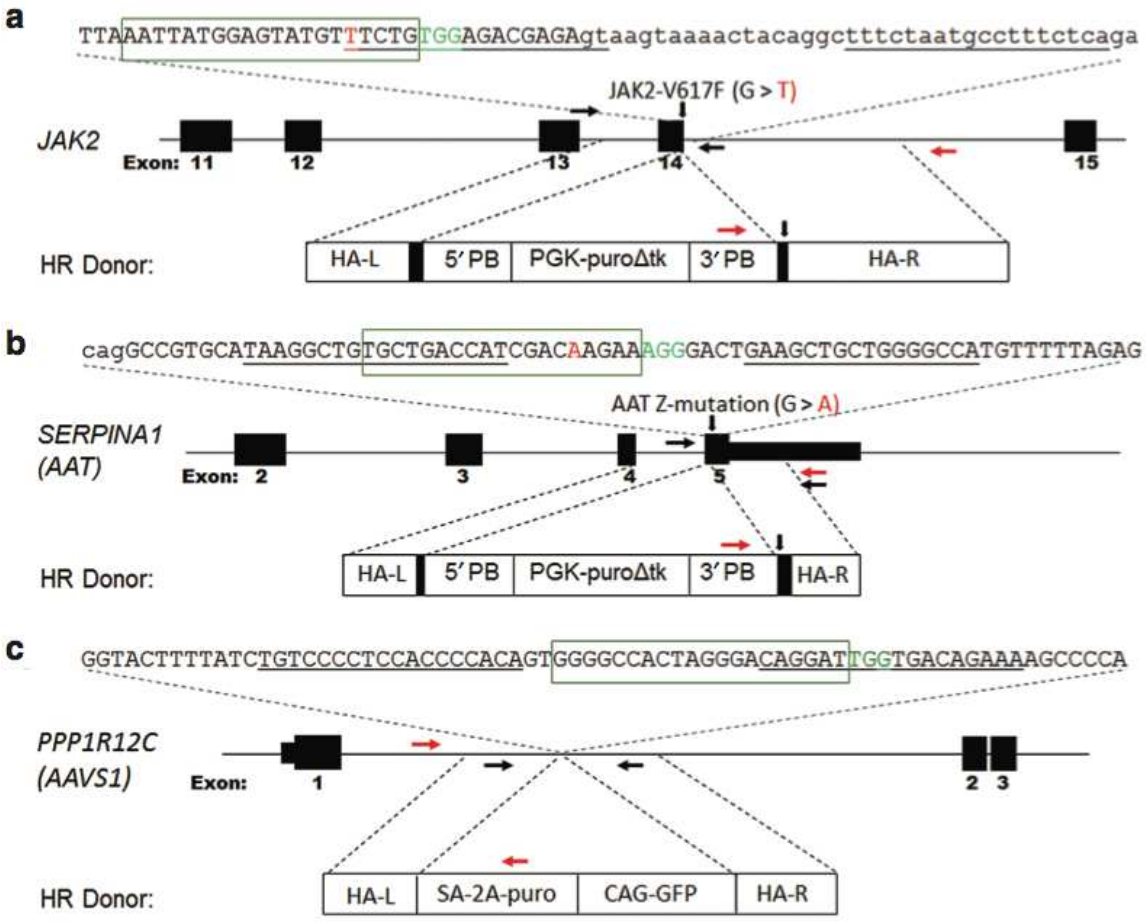
Comparable efficiencies between Cas9-gRNAs and TALENs in facilitating genome editing by homologous donors

One common approach in precise genome editing is to introduce a homology donor together with the engineered endonucleases into the target cells. To assess the ability of Cas9-gRNAs and TALENs in promoting homology-directed repair, we conducted gene targeting experiments with homologous donors at JAK2, SERPINA1, and AAVS1 loci in the same human iPSCs as we did in the MiSeq/NHEJ experiments. We designed an HR donor template with homology arms near the Cas9-gRNA and TALEN cutting sites at the JAK2 locus (Figure 1a). Previously reported HR donors shown to successfully target the AAT Z-mutation by ZFN and TALEN technologies (Yan et al. 2013; Ye et al. 2009) and

the AAVS1 locus (Hockemeyer et al. 2009) were also used (Figure 1b,c). In each targeting experiment, 2×10^6 iPSCs were cotransfected with a homology donor vector and either the Cas9-gRNA or the TALENs. iPSC colony numbers were counted after puromycin selection, and iPSC clones were randomly picked and expanded. Genomic DNA isolated from each clone was used for PCR screening for targeted integration (TI) events. The absolute targeting efficiencies were calculated as "targeted integration events per million input cells" based on (i) the percentage of randomly selected iPSC clones that are positive for TI; (ii) the total colonies after puromycin-selection; and (iii) the total input cells (Table 1-2). In contrast to what was observed in the NHEJ experiments, transfection of Cas9-gRNA and TALENs designed to target each locus resulted in comparable efficiencies; the largest difference was observed at the AAVS1 site targeting where the Cas9-gRNA showed about twofold advantage over TALENs (Table 2).

Figure 1-1 Genomic loci targeted by Cas9-gRNAs and transcription activator like effector nucleases (TALENs) in human induced pluripotent stem cells (iPSCs). (**a**) The genomic structure and nucleotide sequence around JAK2-V617F mutation in *JAK2* gene are shown. Exon sequence is shown in uppercase. The G to T point mutation in exon 14 is indicated. A diagram of the donor vector used in HR-based *JAK2* targeting (Tables 2 and 3) is shown. The vertical arrow indicates the G>T point mutation in JAK2-V617F. The PGK-puroΔtk dual selection cassette was flanked by *piggyBac* (PB) 5′ and 3′ inverted terminal repeats to facilitate potential footprint-free genome editing after PB-transposase-mediated excision. The vectors contain *JAK2* homology arms flanking the putative Cas9-gRNA and TALEN cutting sites. (**b**) The genomic structure and nucleotide sequence around Z-AAT mutation in *SERPINA1* gene. The G to A variant in exon 5 is indicated. The diagram of a previously reported donor vector used in AAT targeting is shown. (**c**) The genomic structure and nucleotide sequence in intron 1 of the *PPP1R12C* gene. The diagram of a previously reported donor vector used in AAT targeting is shown. Recognition sequences of Cas9-gRNA (boxed, PAM sequences shown in green) and TALENs (underlined) are shown at each locus. PCR primers for identifying targeted integration events (red arrows) and or un-targeted allele (black arrows) are shown. HA-L, left homology arm; HA-R, right homology arm.

Figure 1-1 Genomic loci targeted by Cas9-gRNAs and transcription activator like effector nucleases (TALENs) in human induced pluripotent stem cells (iPSCs).



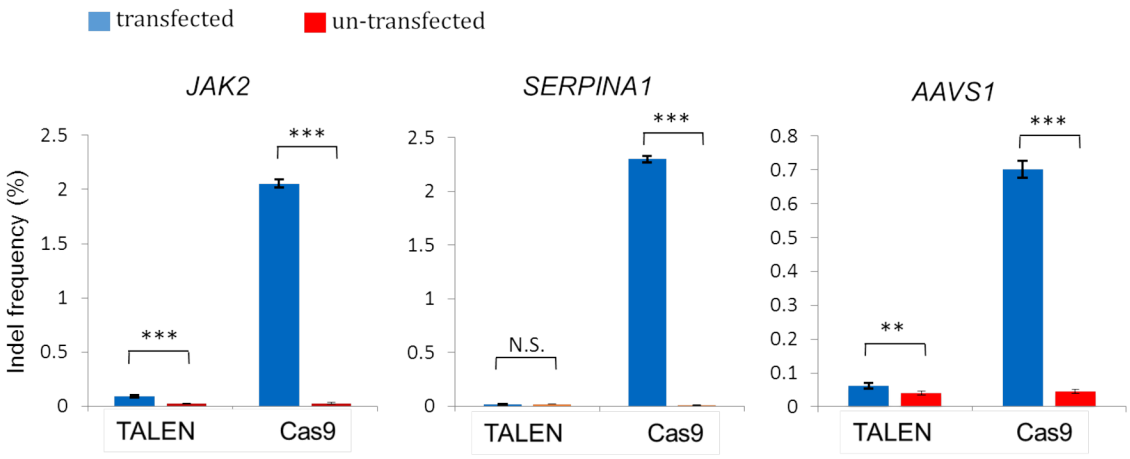


Figure 1-2 Efficiencies of NHEJ-mediated gene disruption at endogenous loci induced by Cas9-gRNAs and transcription activator like effector nucleases (TALENs) in human induced pluripotent stem cells (iPSCs). Frequencies of small indels at indicated locus after Cas9-gRNA or TALEN transfection in human iPSCs. PVB1.4 (homozygous JAK2-V617F mutation), iAAT3 (homozygous AAT Z-mutation) and BC1 (normal control) (**Table 1-1**) were used in experiments targeting *JAK2*, *SERPINA1*, and *AAVS1* loci, respectively. Genomic DNA from un-transfected parental iPSCs was used as background controls (red) compared to those transfected with TALENs or Cas9-gRNA (blue). Error bars indicate 95% Wilson score intervals. N.S., not significant ($P > 0.05$); $^{**}P < 2 \times 10–5$; $^{***}P < 2.2 \times 10–16$. Data represent single transfection and MiSeq experiments.

High specificity of CRISPR/Cas9 in human iPSCs revealed by genome-wide targeted deep sequencing

We next determined the specificity of the gRNAs that together with Cas9 generate indels in human iPSCs at much higher efficiency than TALENs. We also used MiSeq deep sequencing to determine indel frequencies at potential off-target loci with DNA sequences similar to the intended locus. The potential off-target loci were predicted using a previously described bioinformatics program (Hsu et al. 2013), and the top 14–15 loci were measured together with the intended loci (JAK2 and AAVS1) in the same experiments (Figure 1-3). In addition to human iPSCs, we also measured indel frequencies induced by these same Cas9-gRNAs in the widely used human 293T cell line using the same Miseq strategy (Figure 1-3). After transfection with Cas9 and gRNAs, genomic DNA was isolated from the entire transfected population. Upon successful PCR amplification, these loci were analyzed by MiSeq in comparison to those in their parental (un-transfected) cells. Among the 29 most likely off-targets of the two gRNAs, statistically significant indels were detected in 22 of them in 293T cells (Figure 1-3). Two of the most significant AAVS1 off- targets each had >10% of absolute indel rates and were >100-fold higher than observed in the controls (OT-5: 9683/63422 (15.27%) versus 130/401234 (0.03%), $P < 2.2 \times 10\text{-}16$; OT-8: 14261/121156 (11.77%) versus 40/59427 (0.07%), $P < 2.2 \times 10\text{-}16$). In comparison, overall NHEJ rates in iPSCs were ~50-fold lower than those in transfected 293T cells (Figure 1-3). AAVS1 OT-5, with less than 5% of the on-target efficiency, is the only site that showed statistically significant indels in

the targeted iPSCs (Figure 1-4). These results demonstrated the high specificity of Cas9-gRNAs in human iPSCs as compared to 293T cells that have an overall higher level of NHEJ activity and also higher off-target activities by some gRNAs.

Table 2 HDR efficiencies of TALEN and hCas9 at three loci in human iPSCs

Targeted locus	AAVS1[a]		JAK2[b]		AAT[c]	
Nuclease	TALEN	hCas9	TALEN	hCas9	TALEN	hCas9
Total puro colonies[d]	39	74	67	80	146	153
Clones picked	16	22	30	30	18	18
Clones expanded	11	17	27	23	16	15
Targeted integration[e]	11	17	19	15	16	14
Targeted integration/10^6 input iPSCs	13.4	28.6	21.2	20	64.8	59.5

TALEN, transcription activator like effector nuclease.
[a]Experiments performed in BC1 (normal control) iPSC line. [b]Experiments performed in PVB1.4 (PV-specific) iPSC line. [c]Experiments performed in iAAT3 (AAT deficiency-specific) iPSC line. [d]Total numbers of puromycin-resistant colonies from 2×10^6 input cells. [e]As detected by PCR screening. Please see **Figure 1**, **Supplementary Figures S2 and S3** for PCR scheme.

Table 3 Allele-specific gene targeting at point mutation loci by homologous donors in heterozygous human iPSCs

Target point mutation	Guide RNA (nucleotide specificity)	Clones studied	Clones with targeted integration	Integration at both alleles	Integration at one allele	WT allele integration	Mutant allele integration	Alteration at nonintegrated allele
JAK2-V617F[a]	gR-JAK2-F (T)	25	24	0	24	0	24	0
	gR-JAK2-V (G)	29	25	0	25	25	0	0
AAT-Z mutation[b]	gR-AAT-Z (A)	15	15	1	14	1	14	0
	gR-AAT-M (G)	15	14	0	14	14	0	0

[a]Experiments performed in iPV183 (heterozygous JAK2-V617F mutation) iPSC line. [b]Experiments performed in iAAT5 (heterozygous AAT Z-mutation) iPSC line.

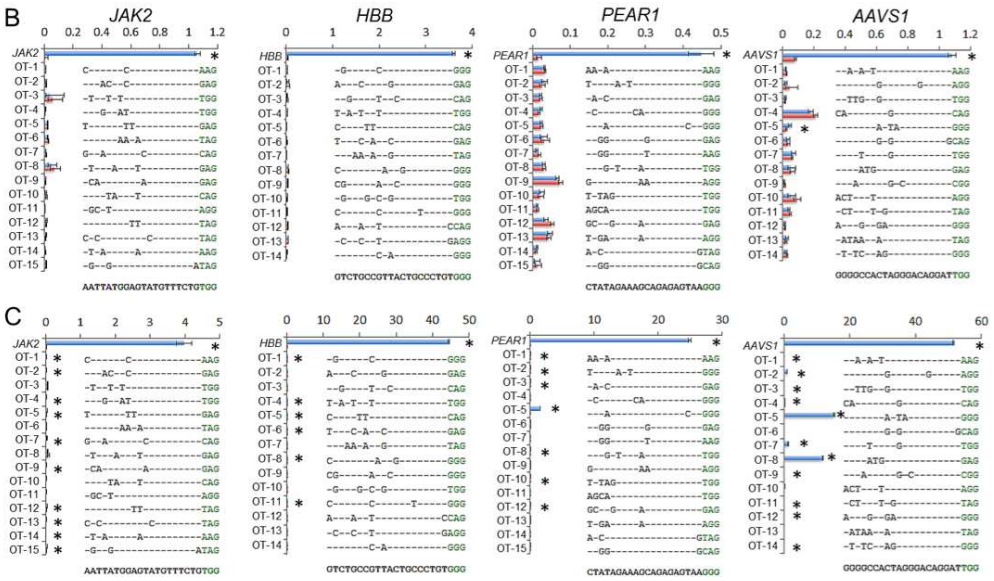


Figure 1-3 Targeted genome-wide deep sequencing of potential off-targets reveals high specificity of CRISPR/Cas9 in human induced pluripotent stem cells (iPSCs). Summarized results of targeted deep sequencing of the 29 bioinformatically predicted most likely off-targets in the human genome. For each Cas9-gRNA, the indel frequencies (x axis) of their intended target and each potential off-target (OT) in the transfected cells (shown in blue) are shown in comparison to the background (DNA from parental cells, shown in red). The mismatched nucleotides of each off-target are shown. The PAM sequence of the target and putative off-targets are shown in green. PVB1.4 (homozygous JAK2-V617F mutation) and BC1 (normal control) iPSC lines were used for *JAK2* and *AAVS1* targeting, respectively. Note that the guide RNA gR-JAK2-F used in this study was designed to target the JAK2-V617F mutation; therefore a single nucleotide mismatch is present in the 293T *JAK2* locus. *Sites with statistically significant ($P < 0.05$) indels above background level. Data represent single transfection and MiSeq experiments.

Allele-specific gene targeting at point mutation loci in human iPSCs using Cas9-gRNAs

Having demonstrated high specificity of Cas9-gRNAs in genome- wide studies, we then investigated the efficiency of targeting specific base pairs in human iPSCs. For this purpose, we again chose to target the somatic JAK2-V617F (rs77375493, G>T) mutation and the inherited α1-antitrypsin (AAT) Z mutation (rs28929474, G>A) in the SERPINA1 gene. Two panels of healthy donor and patient-derived iPSCs carrying either a homozygous wild-type target allele, heterozygous mutant allele, or homozygous mutant allele were used in these experiments (Cheng et al. 2012; Yan et al. 2013; Choi et al. 2011; Ye et al. 2014; Ding et al. 2013) (Table 1-1). Two gRNAs, gRNA-JAK2-F, and gRNA-JAK2-V, were designed with a single nucleotide difference to specifically recognize the V617F mutant allele (T) or the wild-type allele (G) of JAK2 (Figure 1-4a). Another pair of gRNAs, gRNA-AAT-Z and gRNA-AAT-M, were designed to specifically recognize the AAT Z-allele (A) or the wild-type M-allele (G) (Figure 4b). We examined the efficiency and specificity of the two pairs of gRNAs in targeting the point mutations at endogenous loci by cotransfection with Cas9 (Figure 1-3). Indel frequency analyses of the JAK2 gene revealed that only the gRNA-JAK2-F successfully targeted the JAK2 locus in the JAK2-V617F homozygous PVB2.7 and PVB1.4 iPSCs. Likewise, only gRNA-JAK2-V successfully targeted the wild-type BC1 and PVB1.11 iPSCs (Figure 1-4c). The single-nucleotide mismatch between gRNAs and target genomic sequence prevented any major indel mutagenesis by Cas9. Similar results were observed in AAT targeting. The gRNA-AAT-Z, designed to target the Z-allele, had no

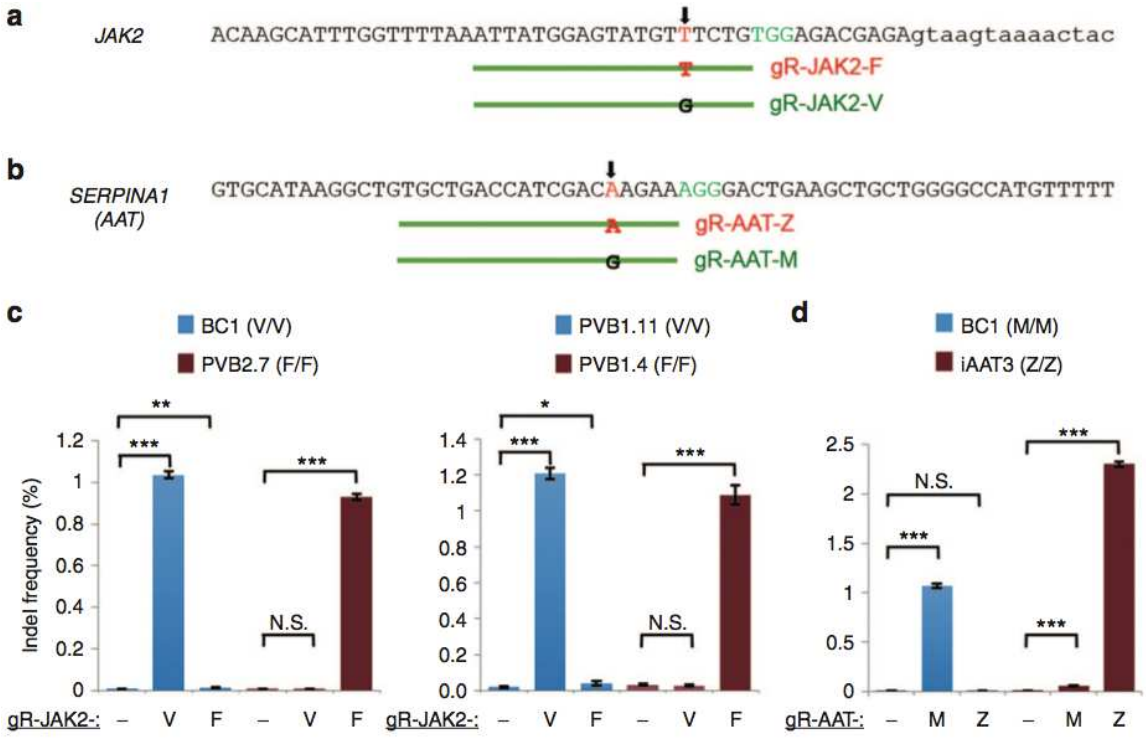
detectable effect on the wild-type AAT allele, while significant indels were observed after transfection into iPSCs that carry the homozygous Z-mutation. Although a statistically significant (compared to the un-transfected control) indel rate was observed at the Z-alleles in iAAT3 (homozygous mutant) after gRNA-AAT- M targeting, this indel rate (0.055%) is 40-fold lower than that observed after gRNA-AAT-Z transfection (Figure 4d).

For gene/cell therapy and disease modeling, it is often highly desirable to achieve precise genome editing based on homologous donors. We therefore evaluated the specificity of Cas9- gRNAs in this setting at the endogenous JAK2 and SERPINA1 loci, using either the PV-iPSC line iPV183 heterozygous for the JAK2-V617F point mutation (Ye et al. 2014) or the iPSC line iAAT5 heterozygous for the Z-AAT point mutation (Choi et al. 2013; Choi et al. 2011). To investigate the allelic specificity of inducing integration events by Cas9-gRNAs, we cotransfected iPV183 with a homology donor template and either Cas9/gRNA-JAK2-F or Cas9/gRNA-JAK2-V, and iAAT5 with donor template and either Cas9/gRNA-AAT-Z or Cas9/ gRNA-AAT-M. Candidate targeted iPSC clones were randomly picked after cotransfection and puromycin selection expanded and screened by PCR and sequencing to confirm targeted integration at the targeted allele as well as the sequence integrity of the nontargeted allele. Among the expanded PV-iPSC clones, 24/25 and 25/29 had targeted integration by Cas9/gRNA-JAK2-F and Cas9/gRNA-JAK2-V, respectively. Among the expanded AAT-iPSC clones, 15/15 and 14/15 had targeted integration by Cas9/gRNA-AAT-Z and Cas9/ gRNA-AAT-M, respectively. Strikingly, in all 49 clones with targeted integration at JAK2 locus, the

integration events occurred only at the JAK2 allele specified by the gRNAs (Table 1-3). Similarly there was only one clone from the Cas9/gRNA-AAT-Z experiment that had targeted integration in both alleles; in all other clones the integration events occurred only at the SERPINA1 allele specified by the gRNAs (Table 1-3). In addition, sequencing of each targeted clone showed the absence of NHEJ-mediated mutations on the other non-targeted allele, further demonstrating the targeting specificity (Table 1-3). Taken together, our results from these patient-specific iPSCs strongly suggest the feasibility of specifically targeting a single-nucleotide mutation or variant by CRISPR/Cas9 in human iPSCs.

Figure 1-4 Allele-specific gene targeting of JAK2-V617F and Z-AAT mutations in patient-specific induced pluripotent stem cells (iPSCs). (**a**) Guide RNAs gR-JAK2-F and gR-JAK2-V were designed to specifically target the mutant allele (with nucleotide T) and wild-type allele (with nucleotide G), respectively. (**b**) Guide RNAs gR-AAT-Z and gR-AAT-M were designed to specifically target the mutant allele (with nucleotide A) and wild-type allele (with nucleotide G), respectively. (**c**) The specificity of gRNA-JAK2-F and gRNA-JAK2-V that differ by a single nucleotide were evaluated in iPSC lines with V/V or F/F genotype. Human iPSCs with WT *JAK2* (BC1 and PVB1.11) or mutant *JAK2* (PVB1.4 and PVB2.7, both are homozygous for JAK2-V617F mutation) were cotransfected with Cas9 and either one of the gRNA constructs. The indel frequency at the *JAK2* locus in each sample is shown. Guide RNA used in each transfection is indicated underneath the indel frequency. (**d**) The specificity of gRNA-AAT-Z and gRNA-AAT-M that differ by a single nucleotide were evaluated in iPSC lines with M/M (control BC1) or Z/Z (AAT deficiency patient-specific iAAT3, homozygous AAT Z-mutation) genotype. Major indel events (>0.1%) were only observed when an iPSC line was transfected with the gRNA designed for its genotype. Error bars in **c** and **d** indicate 95% Wilson score intervals. N.S., not significant ($P > 0.05$); *$P < 0.001$; **$P < 2\times10–5$; ***$P < 2.2\times10–16$.

Figure 1-4 Allele-specific gene targeting of JAK2-V617F and Z-AAT mutations in patient-specific induced pluripotent stem cells (iPSCs).



AIM 1 DISCUSSION

This quantitative comparison of Cas9-gRNAs and TALENs at three endogenous loci in human iPSCs suggests a higher efficiency of gene disruption by the Cas9 endonuclease, corroborating with recent studies (Mali et al. 2013a; Yang et al. 2013b; Ding et al. 2013). Although not statistically significant, the current study also suggests an inverse correlation between gene expression level and the advantage of Cas9 over TALEN; the lowest differential indel frequency was observed at AAVS1, which has the highest level of expression in human iPSCs among the three genes. Conversely, the most advantageous indel induction by Cas9 was observed at the SERPINA1 locus, which has

the lowest expression level (based on RNA-Seq data from ENCODE/Caltech, GSE33480) (Ding et al. 2013). These data demonstrate that the CRISPR/Cas9 system can efficiently target both expressed and nonexpressed loci in human iPSCs, and may have a particular advantage at nonexpressed loci.

In our previous studies, we have observed high efficiency gene targeting at both AAT and AAVS1 loci in human iPSCs using TALENs (Choi et al. 2013; Yan et al. 2013). Therefore, the significantly lower targeting efficiency assayed by indel frequency observed in this study was unexpected. We further examined whether the higher rates of NHEJ-mediated gene disruption correlate with more efficient gene-editing using donor templates with homology arms. We chose to conduct such experiments using the conventional drug selection-based donors as we did in previous studies (Choi et al. 2013; Yan et al. 2013). Although detecting NHEJ and HDR in the same sets of experiments would have been more quantitative, practically this can only be done with a short donor template (such as oligonucleotide donors) in order for both events to be detected by MiSeq sequencing. The current efficiency of targeting by an oligonucleotide donor, however, is still extremely low in human iPSCs on average (Yang et al. 2013b; Soldner et al. 2011). Without drug selection, it is technically difficult to identify targeted clones before breakthroughs in technology are made. Therefore, we focus on the efficiency of the conventional selection-based donor in the current study because it remains a major genome-editing tool for most investigators engaged in research using human iPSCs.

Our quantitative investigation of Cas9 and TALENs in facilitating targeted integration events revealed comparable efficiencies of the two technologies (Table 1-2),

even though their ability to induce small indels at the same endogenous loci varied dramatically (Figure 1-2). There are multiple potential explanations for this. First, the difference in binding/releasing kinetics of TALE or Cas9 to genomic DNA (or other molecules interacting with either Cas9 or TALENs), may differentially affect the subsequent recruitment of the components required for either NHEJ or HDR. As a result, TALENs have a relative preference for HDR over NHEJ at induced DSBs. Alternatively, TALENs generate a higher percentage of single-strand DNA breaks between their binding sites while Cas9- gRNA binding results in more DSBs. Studies using engineered DNA nickases have shown that the single-strand DNA breaks can stimulate efficient HDR without inducing the error-prone NHEJ pathway (Ramirez et al. 2012; Ran et al. 2013; Shen et al. 2014). Finally, TALENs typically generate DSBs with single-strand overhangs in the space between the two TALE-binding sites while Cas9 has been reported to make blunt-end DSBs (Jinek et al. 2012; Maresca et al. 2013). Different types of DSBs generated by TALENs and Cas9 may result in different preference for repair pathway. We anticipate that future studies at the molecular level of nuclease-DNA interaction and in DNA repair pathways may shed light on the exact mechanism, which can aid to further improve this gene targeting technology.

We further examined the specificity of CRISPR/Cas9 system in human iPSCs using targeted deep sequencing, which provides the most accurate measurement of small indels at endogenous loci (Mali et al. 2013b; Cong et al. 2013; Hendel et al. 2014). While a WGS approach offers broad genome coverage with an emphasis on the quality of individual clones, the approach used in this study allowed for a more quantitative analysis

by examining predicted off-targets in significantly greater depth in the entire cell population that has undergone genome editing. Each approach has its own advantages and disadvantages. Importantly, we have now demonstrated the high specificity of CRISPR/Cas9 in human iPSCs using these complementary methodologies.

The observed high specificity and efficiency provide a basis for us to conduct allele-specific gene targeting of point mutations in patient-specific iPSCs. Targeting two disease-associated point mutations, the inherited AAT Z-mutation and the acquired JAK2-V617F mutation, in five human iPSC lines, we have shown that each Cas9- gRNA almost exclusively targets its intended sequence and generates only background levels of indels at the other allele that differs by a single nucleotide alone. Using iPSC lines that are heterozygous for either mutation, we have also shown allele-specific targeting in a setting that is more critical for making precise editing for molecular therapy purposes. An additional advantage of using heterozygous cell lines in this study is that it provides an internal control for each Cas9-gRNA and eliminates the experimental variations such as transfection efficiency that may compromise the results.

We anticipate that allele-specific gene targeting can be widely applied to many other loci in human iPSCs using the CRISPR/Cas9 technology. It should be noted that one limitation of the current technique is that it will require a PAM sequence close enough to the variant of interest. In this study, each point mutation is located at the 5th nucleotide 5′ of their respective PAM sequence. For other genetic variants that locate further from PAM sequences, the specificity of Cas9-gRNAs may decrease. However, for certain mutations such as the JAK2-V617F and Z-AAT reported here, it offers a unique

targeting specificity that was not easily achieved by previous technologies. Additionally, as new PAM sequences become targetable by adapting Cas9 from other species and/or through protein engineering, we anticipate that this technique will be applicable to more loci in the near future.

Aim 2: Investigate the genomic integrity of human iPSCs after genome editing by CRISPR/Cas9 or TALENs.

The ability to precisely modify DNA in human iPSCs has been greatly improved by targeted nuclease technologies including CRISPR/Cas9, and TALENs but their specificity towards the rest of the genome remains unclear. After initial studies reported higher than expected off-target mutagenesis using cancer cell lines, a deeper investigation into their specificity was warranted in human iPSCs before further applications in disease modeling or regenerative medicine could be applied. Several targeted sequencing, or Cas9 capture approaches have been conducted to search for off-target mutagenesis but an un-biased genome wide screen was needed to know the full extent of genomic mutations acquired during the genome editing process. To this end whole genome sequencing was conducted on four iPSC clones derived from a single healthy donor targeted with an AAVS1 knock-in stimulated by either TALENs, or CRISPR/Cas9 and compared to DNA from parental iPSCs harvested directly before transfection. ~200 SNVs and ~10 indels were identified in each clone but none of the mutations were recurrent or similar to the nuclease binding site; implying that the mutations were either random or pre-existing in the original population and not a result of the genome editing process. These results from whole-genome sequencing (WGS) analysis of CRISPR/Cas9 and TALEN-targeted human iPSC clones demonstrate that these engineered endonucleases provide efficient genome-editing tools with high specificity.

AIM 2 INTRODUCTION

Human iPSCs provide renewable cell sources for human biology and disease research and the potential for developing gene and cell therapy. Known concerns of chromosomal instability and mutagenesis after prolonged culture of human stem cells lead to in depth investigation of the genomic integrity after cellular reprogramming to generate iPSCs. Initial sequencing of reprogrammed cell lines reported ~6 exonic mutations but alarmingly there was a disproportionate number of cancer associated genes (Gore et al. 2011). Additional studies confirmed the previously mentioned mutation rate but did not find any rise in mutation among genes associated with cancer (Cheng et al. 2012). However the full utilization of iPSCs requires the ability to precisely edit their genome in a targeted manor without the induction of undesired off-target mutations.

Realization of this potential will rely in part on our ability to precisely edit or engineer the human genome in an efficient way. Recent developments in designer endonuclease technologies such as ZFN, TALEN, and CRISPR/Cas9 endonuclease have provided ways to significantly improve genome editing efficiency in human iPSCs. Some analyses using cancer cell lines reported higher-than-expected levels of off-target mutagenesis by Cas9-gRNAs (Fu, 2013; Hsu, 2013), raising concerns about the practical applicability of this approach in therapeutic contexts. Some recent studies, including one on human adult stem cells, showed a minimal level of off-target effects by CRISPR/Cas9 (Schwank, 2013). However, these existing analyses of off-target effects and mutational load in gene-corrected stem cells have been restricted to checking predicted off target sites and are therefore limited in scope. To assess the value of this type of gene editing

approach for therapeutic applications, it is critical to rigorously examine whether it is possible to generate gene-edited cell lines with minimal mutational load.

To investigate this we have conducted whole-genome sequencing of four iPSC clones successfully targeted at the AAVS1 locus, a ‘‘safe harbor’’ in the human genome that is used for stable transgene expression in a variety of contexts. To generate the lines, we used an integration-free human iPSC line, BC1, whose genomic integrity has been characterized in detail by next-generation sequencing (Cheng, 2012) and targeted a GFP expression cassette into the AAVS1 site with either a previously reported Cas9-gRNA combination or a pair of improved heterodimeric TALENs (Mali, 2013; Yan, 2013). Twenty days after transfection of the donor plasmid and either the TALENs or Cas9-gRNA into BC1, we harvested four clones with confirmed targeted integration (hCas9-C4, hCas9-C16, TALEN-C3, and TALEN-C6; Table 2-2) and the parental BC1 iPSCs for whole-genome sequencing.

AIM 2 METHODS

Maintenance and expansion of human iPSCs

Human iPSCs were cultured with E8 medium (Life Technologies) on tissue culture plates coated with Matrigel (BD Biosciences) or Vitronectin (Life Technologies). For routine passaging, iPSCs were digested with Accutase (Sigma) for 5 minutes and washed with PBS by centrifugation at 200xg for 5 minutes. Digested iPSCs were then plated at a density of $2x10^4$ per cm^2 with E8 medium supplemented with 10 uM ROCK Inhibitor Y-27632 (Stemgent) for the first 24 hours.

HR-mediated gene targeting at the AAVS1 locus in human iPSCs

2 million BC1 iPSCs were nucleofected with either: 1) 5μg AAV-CAGGS-EGFP (addgene # 22212), 3μg each of a heterodimeric TALEN pair targeting the AAVS1 (Yan 2013). 2) 5μg AAV-CAGGS- EGFP (Addgene # 22212), 3μg hCas9 (Addgene #41815), and 3μg AAVS1-T2 gRNA (Addgene #41815). Nucleofected cells were plated on 2 Matrigel coated 6-well plates and were subjected to puromycin selection (0.5 μg/mL) from day four to 11 at which point individual colonies were picked and expanded. At day 20 genomic DNA was isolated and screened for targeted integration (TI) by PCR and Sanger sequencing. Clones were also screened for the presence of a wild type allele without TI by a PCR amplification using a 30 second extension time insufficient to amplify the 3.8-kb TI alleles.

WGS and mutation calling in TALEN vs Cas9/gRNA targeted iPSCs

Among clones with confirmed targeted integration and lack of random integration of donor vector, two were randomly selected for both TALEN and Cas9/gRNA targeted iPSCs, and genomic DNA from these clones and the BC1 parental cell line was prepared for Illumina sequencing. Sequencing library construction was performed as previously described (Gore 2011). Briefly, approximately 1.5 – 3 μg of genomic DNA was purified from cells of each sample and sheared with a Covaris AFA. The DNA fragments were then end-repaired, A-tailed, and ligated to Illumina barcoded sequencing adaptors. The ligated products were amplified by PCR to generate barcoded whole-genome sequencing libraries. Each library was pooled together, and the pool was sequenced using two

Illumina HiSeq 2000 flowcells (with a variable number of sequencing reads for each library based on pooling ratio).

In order to identify candidate mutations in each edited cell line, variant calling was performed as previously described (Gore 2011). Briefly, reads from each flowcell were de-barcoded and assigned to each cell line. Reads passing Illumina's chastity filter were mapped to the human reference genome (hg19) using BWA. Clonal duplicate reads caused by PCR were then removed using Picard- tools. The remaining nonclonal reads were processed as per the GATK Best Practices for small sample sizes, including realignment about indels and quality score recalibration. Both SNVs and indel variants were then called for each sample using the GATK HaplotypeCaller, with filtering parameters based on the GATK Best Practices; additionally, sequencing reads were discarded if a single end contributed to more than one SNV or indel. Sites where each DNA-edited cell line showed either a heterozygous SNP or an indel call that was not observed in the unedited cell line were considered as mutations if at least 30% of reads in the edited cell line contained the mutation, no allelic content derived from the mutation was present in the unedited line, and if the candidate mutation had not previously been observed in the dbSNP database or other unrelated samples.

AIM 2 RESULTS

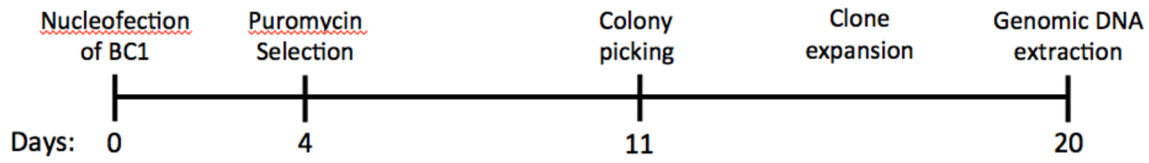
CRISPR/Cas9 or TALEN stimulated KI to the AAVS1 safe harbor locus

When comparing total number of puromycin resistant colonies CRISPR/Cas9 had a two-fold advantage but both were equivalent in their 100% targeting efficiency, of all 28 puromycin resistant iPSC clones screened had the intended AAVS1 KI by PCR. Candidates for targeted integration were then confirmed by Sanger sequencing of the 5' and 3' juntions. iPSC clones with single allele AAVS1 KI were screened for by the presence of untargeted wild type band using a short extension time to avoid the 4.5kb product from the targeted integration. TALEN and Cas9 were equivalent in the rate of bi-allelic integration at about ~33% of clones analyzed. Among the majority of clones with single allele integration the other allele was sequenced to screen for indel mutations caused by NHEJ. None of the TALEN single-allele targeted clones (0 of 6) contained a mutation on the other allele while 50% (3 of 6) Cas9 targeted mutations contained indels. Although all the screened iPSC clones had the targeted AAVS1 KI but additional random integration of the donor vector at other genomic loci was screened for using a PCR strategy unique to the donor vector but outside the homology arms and intended integration. Random integration was high in both Cas9 and TALEN at ~65% removing most candidate clones from further characterization. Combining all these screening criteria two top candidates for each TALEN and Cas9 targeted clones were selected along with parental BC1 iPSCs for WGS.

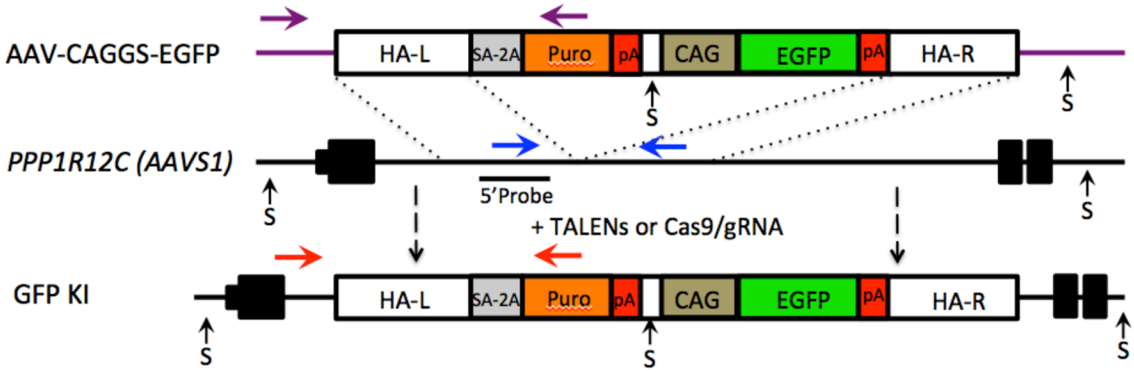
Figure 2-1 AAVS1 targeting diagram, timeline, and screening strategy. (**a**) Timeline of gene targeting in BC1 iPSCs and genomic DNA isolation for sequencing. BC1 iPSCs were co-nucleofected by a donor plasmid AAV-CAGGS-EGFP with either TALENs or Cas9-gRNA. Nucleofected cells were cultured at a low density for four days before puromycin selection. Individual puromycin-resistant clones picked at day 11 were further expanded for additional nine days. Genomic DNAs isolated at day-20 were used for analyses of targeted integration event and for whole genome sequencing. (**b**) Diagrams of the donor plasmid AAV-CAGGS-EGFP, the native AAVS1 (*PPP1R12C*) locus structure and the genomic structure after targeted integration of GFP expression cassette (GFP KI). Primers for PCR amplification of targeted allele (red), untargeted allele (blue) and random integration (purple) were shown. DNA probe (5' probe) for southern analysis of SphI (S) digested genomic DNA were also indicated. The HR-mediated targeted integration at AAVS1 by TALENs and Cas9-gRNA. $2x10^6$ BC1 iPSCs were used for each experiment. Puromycin-resistant iPSC clones were randomly picked, expanded and screened by PCR to identify the HR events leading to targeted integration (TI).

Figure 2-1 AAVS1 targeting diagram, timeline, and screening strategy.

(a)



(b)



WGS of CRISPR/Cas9 or TALEN targeted iPSC clones

The sequencing reads, ranging from 83 Gbps to 100 Gbps from each targeted clone, were first aligned to the human hg19 reference genome to enable identification of single-nucleotide variants (SNVs) and small indels (Table 2-1). Our analysis identified ≥4.2 million SNVs and ≥500,000 indels in each genome (Table 2-2) in comparison to the hg19 reference genome, suggesting that it is a rigorous data set that covers the genome in sufficient depth to detect sequence variants. The ‘‘germ-line’’ variants (present in BC1 parental iPSCs and different from hg19) were readily detectable in each targeted cell line (80%–88%), indicating that the sensitivity of variant detection in our analysis is high (Table 2-2). The variations from each targeted clone were then compared to the BC1

parental iPSCs to enable the generation of a list of potential variations arising during the gene editing process, which we then confirmed using genomic PCR and Sanger sequencing. We confirmed 62 out of 69 SNVs tested for an overall confirmation rate of 90%, and based on that we estimate that the total SNVs in the four iPSC clones range between 217 and 281 and that the total indels range between 7 and 12 (Table 2-2). Overall the genomic variation levels in TALEN- and Cas9-targeted groups were comparable.

No SNVs or indels were recurrent or similar their respective nuclease binding sites.

One important consideration is how many of these detected SNVs and indels were the results of off-target mutagenesis by the engineered endonucleases. To address this question, we generated a list of 3,665 (Cas9) and 238 (TALEN) putative off-target positions by using the EMBOSS fuzznuc software package. Each candidate SNV and indel was compared to this list and none of them are within a potential off-target region (Table 2-2), consistent with previous analyses looking at predicted off-target sites. Our analysis also shows that each SNV and indel is unique and that none of them occurred in more than one cell line. The absence of recurring mutations and the fact that none of the mutations resides in any putative off-target site by bioinformatic prediction strongly suggest that these mutations were randomly accumulated during regular cell expansion and are not direct results of off-target activities by Cas9 or TALENs.

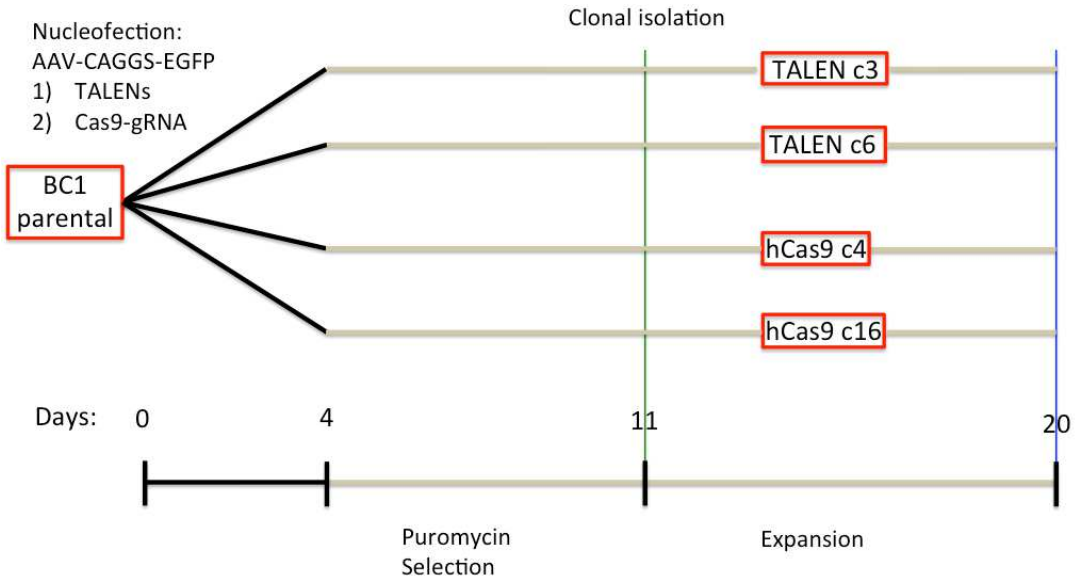


Figure 2-2 Genome editing timeline for clonal isolation of targeted iPSCs. BC1 parental iPSCs were transfected with AAV-CAGGS-EGFP (HDR donor vector) and either: 1) AAVS1-TALEN-L + AAVS1-TALEN-R or 2) Cas9 + gRNA-AAVS1-T2 and plated at low density (2 x 6 well plates) for clonal isolation. Puromycin selection was started at day 4 and continued throughout the clonal isolation process. At day 11 colonies were manually picked and expanded individually for genomic DNA extraction and continued culture.

Table 2-1 WGS summary of TALEN vs Cas9 targeted iPSCs

iPSC line	Sequence Generated	Genome Coverage (fold)	Usable Read Rate	Total SNVs (compared to BC1)	Indels (compared to BC1)	Variations in predicted off-targets	Recurring variations
BC1 (parental)	77 Gbps	26x	49%	N/A	N/A	N/A	N/A
hCas9-C4	100 Gbps	33x	62%	217	7	0	0
hCas9-C16	95 Gbps	32x	55%	281	9	0	0
TALEN-C3	83 Gbps	28x	59%	230	12	0	0
TALEN-C6	95 Gbps	32x	59%	221	7	0	0

Table 2-2 Whole-Genome Sequencing Variant Details.

Sample	# SNVs Called (Compared to hg19)	% BC1 Germline SNVs Detected	# Indels Called (Compared to hg19)	% BC1 Germline Indels Detected
BC1 (parental)	4,203,209	N/A	509,184	N/A
hCas9-C4	4,364,174	88.3%	588,965	78.8%
hCas9-C16	4,316,078	79.7%	581,827	77.0%
TALEN-C3	4,287,383	84.7%	550,262	74.6%
TALEN-C6	4,322,626	80.0%	583,529	77.2%

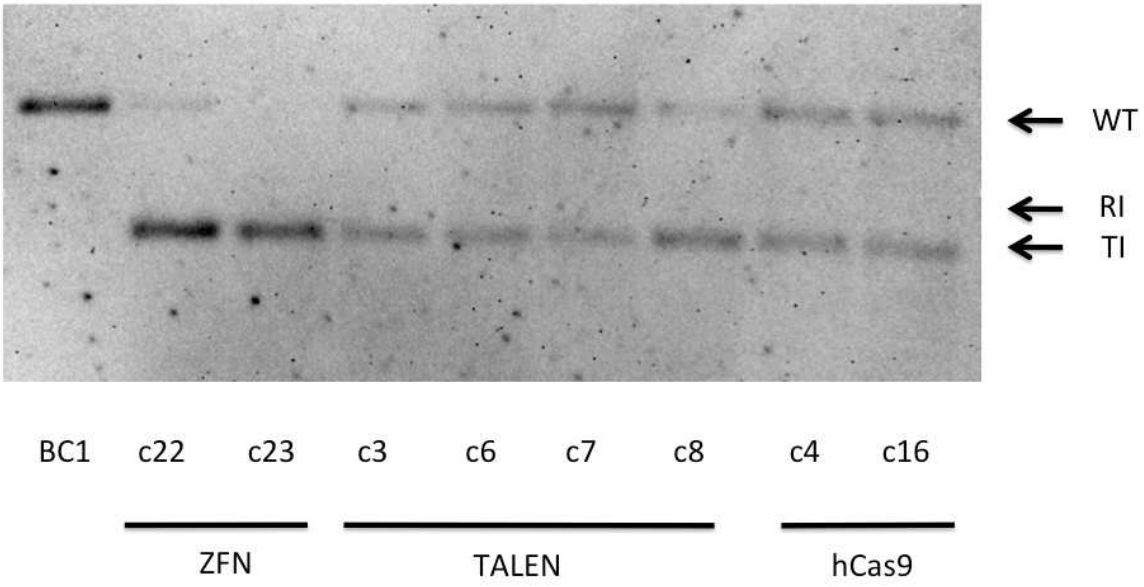


Figure 2-3 Southern blot confirms targeted integration into the AAVS1 safe harbor locus. Southern blot analysis was performed on selected iPSC clones that were positive for a targeted event (TI) and were negative for random integration based on PCR screenings. Using the DNA probe shown in Figure 2-1a, hybridization to DNA with targeted integration would result in 3.8 kb band while un-targeted DNA would result in 6.5 kb band. Four targeted iPSC clones (hCas9-C4, C16 and TALEN-C3, C6) were selected for whole genome sequencing in comparison to parental BC1 iPSCs.

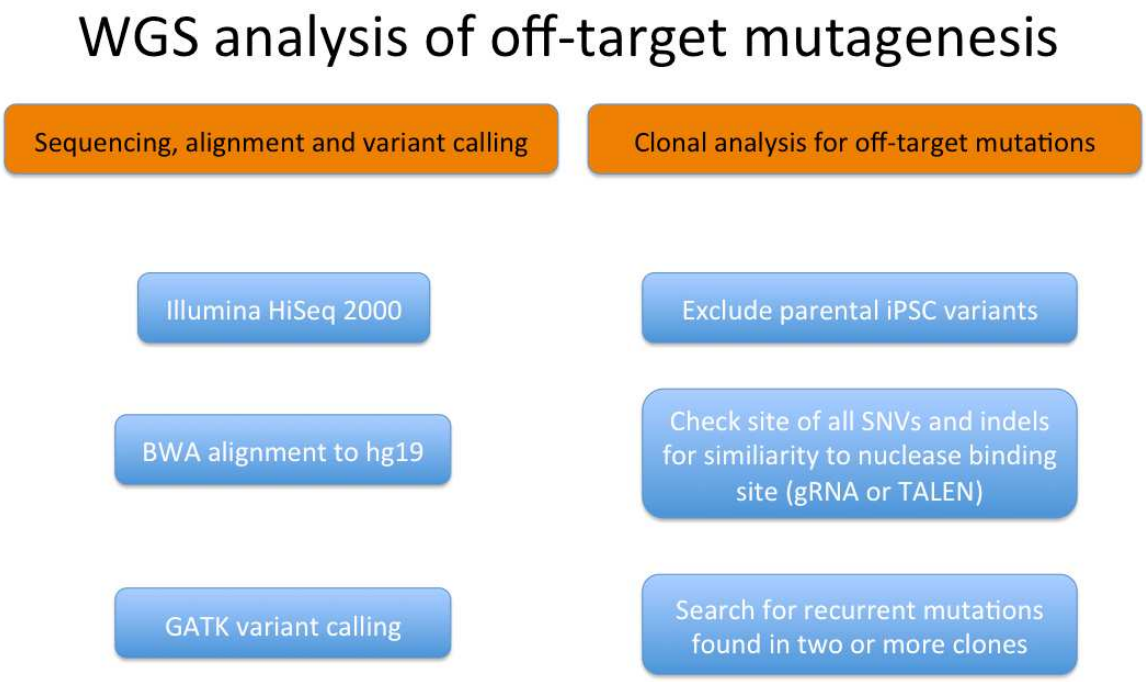


Figure 2-4 WGS alignment, variant calling, and analysis pipeline screening for off-target mutagenesis. After confirming 2 TALEN and 2 Cas9 successfully edited iPSCs genomic DNA was used for sequencing using the illumina HiSeq 2000 to generate ~30x coverage. These reads were then aligned to the reference genome (hg19) using BWA and variants were called using GATK with local re-alignment for indel analysis. To investigate mutations acquired during the genome editing process all variants found in the parental BC1 iPSCs were excluded from each of the targeted clones. All new mutations were screened for similarity to the nuclease binding site (within 100 bps) up to 5 mismatches from the gRNA or TALEN sequence. Recurrent mutations were also searched for among all the targeted clones for signs of rare variants found in the parental iPSCs or a non-sequence based off-target site.

AIM 2 DISCUSSION

To fully utilize iPSCs for disease modeling or regenerative medicine requires the ability to precisely modify these cells without inducing undesired off-target mutagenesis that would confound results or could lead to oncogenic transformation. Many approaches have been taken by others such as targeted amplicons sequencing of predicted off-target loci, or target capture to enrich for DNA at these sites but an unbiased approach to search genome wide for mutagenesis was lacking to fully investigate the genomic integrity after gene targeting combined with designer nucleases. When comparing overall efficiencies of targeted integration at the AAVS1 locus Cas9 had a two fold advantage of TALEN but both generated more colonies than could be picked. For this experiment only single allele targeted clones were selected for further analysis and the untargeted allele was screened for indels caused by NHEJ from the nuclease; confirming previous work in aim one Cas9 leads to more frequent disruption of the other allele, found in 50% of clones, while TALEN targeted clones were all wild-type at the untargeted allele.

In the previous aim based on previous deep sequencing of predicted off-target loci conducted in aim one only a single locus showed significant off-target cleavage (OT-5, chr13:106612912) in the bulk transfected cells but was only slightly elevated over the background mutation rate; this site was not mutated in any of the sequenced clones. All sites in the genome within five mismatches to the nuclease binding site were also screened but no mutations were identified in all clones screened indicating that for these gRNA and TALEN genome edited clones can readily be derived with no evidence for off-target mutagenesis at sites of similar sequence.

These results from whole-genome sequencing analysis of Cas9- and TALEN-targeted human iPSC clones demonstrate that these engineered endonucleases provide efficient genome-editing tools with high specificity. It remains to be clarified whether the higher off-target rates observed in cancer cell lines are due to the overexpression of gRNAs and Cas9 protein and/or due to exacerbated and faulty DNA repair in these cell types. The higher specificity observed in human iPSCs, combined with the rapid development of next-generation sequencing technology, makes it possible to characterize and isolate high quality genome-edited stem cell clones with minimal mutational load. The guiding principle established with human iPSCs will likely be applicable to other types of stem cells and come with improvements in gene transfer and targeting efficiencies. Our current study of gene targeting in human iPSCs will help to establish better models for human biology and disease research and to provide proof-of-principle for future gene therapy.

Aim 3: Knockout the PEAR1 gene and its cis-regulatory element to evaluate their function and regulation during iPSC-derived megakaryopoiesis.

ABSTRACT

Genetic variation within platelet endothelial aggregation receptor 1 (*PEAR1*) has been associated with platelet aggregation and was later reported to act as a negative regulator of cell proliferation and lineage commitment during megakaryopoiesis. To investigate this a *PEAR1* KO iPSC line was generated that resulted in a homozygous frameshift mutation that ablated PEAR1 expression during iPSC derived megakaryocyte differentiation. When compared to otherwise isogenic iPSCs, *PEAR1* KO iPSCs displayed accelerated megakaryocyte lineage commitment during hematopoietic differentiation as measured by cell surface expression of CD41 and CD42 which was accompanied by an increased cell proliferation which continued through megakaryocyte differentiation. PEAR1 cell surface expression naturally increased over the course of hematopoietic and megakaryocyte differentiation but these results indicate that it is not essential for their proper maturation and acts as a negative regulator of cell growth during the process. A SNP (rs12041331) within the first intron of *PEAR1* was reported to reside within an enhancer by a luciferase assay while more investigation is required to elucidate its role in the native chromatin context of an appropriate cell type. BC1 cis*PEAR1* KO was created with a 251 bp single allele deletion containing rs12041331, leaving the other

allele unmodified by Cas9. When this clone was compared to parental BC1 iPSCs *PEAR1* mRNA expression was reduced ~50% during hematopoietic differentiation and ~30% during megakaryocyte differentiation demonstrating regulatory effects of this region on gene expression. Further analysis of BC1 cis*PEAR1* KO mRNA revealed that the reduction in expression was primarily on the same allele as the deletion, indicating a cis-regulatory mechanism of action for the region containing rs12041331.

AIM 3 INTRODUCTION

Proper production, activation, and aggregation of platelets play a critical role in normal human hemostasis as well as many states of disease including arterial thrombosis, atherosclerosis and thrombocytopenia. The production of platelets begins in the bone marrow where hematopoietic stem cells divide asymmetrically leading some of their progeny towards the common lymphoid progenitor (CLP) destined for lymphocytes and NK cells, while others are directed towards the common myeloid progenitor (CMP) which gives rise to granulocytes, monocytes, erythrocytes, and platelets. The CMP further differentiates to a megakaryocyte/erythroid progenitor (MEP) that remains bipotent with the capacity for red blood cells and platelets until TPO directs their commitment towards the megakaryocyte progenitor which undergoes a complex process of membrane restructuring, protein production and packaging coupled with endomitosis, a process of nuclear division without cell division, resulting in large polynucleated megakaryocytes that shear into proplatelets and are ultimately released as mature platelets from the vascular sinusoids in the bone marrow into the blood stream. Mature

platelets lack a nucleus but they retain their endoplasmic reticulum which aids in maintaining an inhibited resting state until exogenous agonists such as ADP, collagen or thrombin are detected and lead to activation. Integrin αIIbβ3, also known as glycoprotein IIb/IIIa, is a key membrane bound receptor that is tightly repressed during normal platelet function. Platelet activation leads to a conformational change in Integrin αIIbβ3 exposing its receptor domain allowing fibrinogen or von Willebrand factor to bind, causing sustained platelet-platelet contacts from crosslinking and coagulation and the formation of platelet plugs. Activated platelets create a positive feedback loop recruiting and activating adjacent platelets by releasing ADP, fibrinogen and other factors such as stored α-granules and dense-granules upon activation.

Platelet Endothelial Aggregation Receptor 1 (PEAR1) was initially identified during a proteomic analysis of platelet proteins that undergo post-translational modification in response to platelet activation. Protein domain analysis revealed PEAR1 contains 15 extracellular epidermal like growth factor (EGF) repeats, an EMI protein-protein interaction domain and an intracellular tyrosine (Tyr-925) that becomes phosphorylated upon platelet contact by centrifugation, or by activation by delivery of an agonist which could be blocked by an inhibitor of the integrin αIIbβ3 pathway (Nanda et al., 2005). Further tissue expression analysis by the same group revealed that in addition to platelets, PEAR1 is also expressed in endothelial cells, megakaryocytes, erythroid cells, and at low levels in several other cells types. At this point it is clear that PEAR1 plays a role in signal transduction of platelets and/or endothelial cells but the function of PEAR1 phosphorylation was still unknown at this time.

A GWAS investigating platelet aggregation phenotype identified a PEAR1 intronic SNP, rs12041331 (A/G) that was associated with platelet aggregation phenotype as measured by agonist induction using collage, ADP, or epinephrine. The major variant G, was associated with greater platelet aggregation as well as increased protein content in blood derived platelets (Faraday et al. 2011). To address how intronic genetic variation may influence protein levels rs12041331 was tested for cis-regulatory potential by luciferase assay and found to act as an enhancer while the G variant construct increased expression significantly greater than the A variant. Another independent GWAS linked rs12041331 to platelet aggregation phenotype again with the major variant G associated with greater platelet aggregation, but surprisingly the A variants was associated with increased risk of adverse cardiovascular disease outcomes (Lewis et al. 2013). Normally increased platelet aggregation is a risk factor for adverse clinical outcomes including heart attack and stroke but rs12041331 seems to have opposite effects with regards to these two phenotypes. These paradoxical findings promote further investigation into the role of PEAR1 function in megakaryopoiesis and its potential regulation by genetic variation within the first intron.

Functional characterization revealed that the extracellular domain of PEAR1 binds platelets while the intracellular domain becomes phosphorylated and signals through the PI3K/AKT pathway that leads to activation of integrin αIIbβ3 revealing a functional mechanism to explain the previously reported GWAS hit (Kauskot et al. 2012). The initial studies implicated a functional biological and molecular role for this protein at the final platelet stage but experiments evaluating megakaryopoiesis revealed a

functional role of PEAR1 in cell lineage commitment and growth. When PEAR1 was knocked down using siRNA in human $CD34^+$ cells purified from peripheral blood it resulted in increased megakaryocyte lineage commitment as observed by increased numbers of CFU-MK without influencing CFU-E. Upon in vitro megakaryocyte differentiation PEAR1 knockdown increased cell proliferation but did not inhibit final maturation characteristics including cell surface expression of CD41 and CD42 double positive cells, and production of 4N and 8N polyploidy cells, concluding that it is an attenuator of megakaryopoiesis without disrupting the production of mature megakaryocytes (Kauskot et al. 2013).

This stage of development is better suited for disease modeling in iPSCs as there are defined protocols for hematopoietic and megakaryocyte differentiation while defined techniques for iPSC-derived functional platelets are lacking. In this study CRISPR/Cas9 was used to KO PEAR1 in health iPSCs to compare its effect on megakaryopoiesis when differentiated alongside the parental iPSCs they were derived from. Megakaryocyte development was monitored by expression of integrin, CD41, and glycoprotein 1b, CD42 during the differentiation process. To investigate the regulation and timing of *PEAR1* expression during hematopoiesis and megakaryopoiesis rs12041331 was deleted using two gRNAs spaced 251 bp apart flanking this SNP to disrupt the putative enhancer in which it resides. BC1 cis*PEAR1* KO iPSC line was isolated with a single allele targeted deletion and compared to otherwise isogenic BC1 parental iPSCs by qRT-PCR for relative PEAR1 mRNA analysis. Cis-regulatory effects were isolated by sequencing exon 8 of PEAR1 mRNA, which contains an unrelated SNP, rs7723503 (A/C), used quantify

the relative expression from the wild type allele to the allele containing the intronic deletion. Initial reports that genetic variation within PEAR1 is associated with platelet aggregation phenotype as well as adverse clinical outcomes warrants further investigation into its role and regulation during megakaryopoiesis using genome editing in iPSCs to model these PEAR1 alterations in isolation.

AIM 3 METHODS

Human iPSC based hematopoietic and megakaryocyte differentiation

BC1 iPSCs were grown on Vitronectin coated plated with Essential 8 media media and passaged by accutase for single cell digestion. Spin-Embryoid Body (EB) hematopoietic differentiation followed by five days of megakaryocyte differentiation was carried out as previously described (Liu et al. 2015) to compare genome edited lines to the parental BC1 iPSCs from which they were derived. At day 12 a 70 μm filter was applied to remove the bulk EB cells from the suspension cells containing the hematopoietic stem and progenitor cells (HSPCs) which were counted using the countess to determine total cell number and divide 3 x 10^5 cells for each sample of FACS, 1 x 10^6 for RNA extraction, and 3 x 10^6 cells for continued megakaryocyte differentiation. FACS analysis was conducted for hematopoeitc markers (CD34 and CD45), megakaryocyte markers (CD41 and CD42) and for PEAR1 expression. After five days of megakaryocyte differentiation the cells were analyzed again by FACS and a pellet was saved for RNA expression analysis.

Genome editing PEAR1: coding sequence (Exon 6) and rs12041331 KO

Two gRNAs (P1E6a and P1E6b) were designed to target PEAR1 exon 6 because this exon was present in also isoforms past their transcription start site. They were spaced 173 bp apart to increase the chance of an out of frame deletion as the common breakpoint occurs three nucleotides upstream of the PAM. To delete rs12041331 several combinations of gRNAs were designed and tested in HEK293T cells but only the most efficient pair, gRNA-Dg1 and gRNA-P1g1A, were used to target BC1 iPSCs. 2 million BC1 iPSCs were nucleofected as previously described (Smith et al. 2015) with 3 μg of each of gRNA and 9 μg of pCas9_GFP. Cells were enriched for pCas9_GFP+ cells by FACS and plated at low density (50-200 cells per 6-well) for clonal isolation and screening of PEAR1 target. An additional round of low density seeding and colony picking was repeated to ensure that cells were clonal.

mRNA extraction, cDNA conversion and RT-PCR

In parallel with FACS analysis and DNA isolation to genotype ensure the samples, RNA was extracted from ~1 million cells using the Zymo RNA isolation Kit which removes genomic DNA in two steps including a filter and enzymatic DNase removal. 300 ng – 1μg of total RNA was used as a template for cDNA synthesis was conducted with the Superscript III (Life Technologies) using random hexamers by the manufacturers recommended protocol. RT-PCR for PEAR1 was conducted using the following primers: E8P1F (5'-CAAAATGGAGGTGTCTTCCAA) and E8P1R (5'-ACCGATCCCCAGTGTAACC) using the SYBR green master mix using CT method

and normalizing each sample to GAPDH. All samples were further normalized to the average of BC1 iPSC derived HSPCs over three biological replicates to compare the MK and bulk EB time points.

Deep sequencing for allele specific analysis of PEAR1 exon 8

Starting with cDNA from each cell line (BC1 and BC1 cis*PEAR1* KO) and time point (EB day 12, MK day 5) DNA barcoded primers were used to amplify PEAR1 exon 7 and 8 containing a SNP rs77235035 (A/C) for three biological replicates of each sample conducted at independent differentiation experiments. PCR amplicons were purified using the QIAGEN PCR purification kit and quantified using Qubit dsDNA BR Assay kit. 125 ng of each sample was pooled together, mixed for KAPA library preparation kit (Product # KK8234) for deep sequencing on the MiSeq v2 2x250 sequencing chemistry. After confirming quality, length and throughput of reads using FastQC, reads were demultiplexed using the ea-utils fastq-multx function from both the forward and reverse strands. Each sample was the aligned to the reference sequence for the spliced form of PEAR1 exon 7 and 8 and extracted the total nucleotide counts for the position of rs77235035 to compare allelic mRNA expression. The percentage of reads for each allele were calculated and plotted in Figure 3-12.

AIM 3 RESULTS

An iPSC based model to generate PEAR1 expressing megakaryocytes.

BC1 iPSCs were used as a normal control to test if our iPSC based model to produce HSPCs and megakaryocytes for PEAR1 expression as well as megakaryocyte lineage markers CD41 and CD42. At day 12 of spin-EB hematopoietic differentiation PEAR1 surface expression was detected on ~25% of cells and CD41+CD42+ double positive cells made up about ~20% of the population (Figure 3-1) indicating that a population of megakaryocyte precursors already exists at this stage before specific megakaryocyte induction. Megakaryocyte differentiation was continued for five days with PEAR1 expression reaching 80% while CD41+CD42+ cells reached ~75% indicating a successful megakaryocyte maturation that could be used to further interrogate PEAR1 function from this iPSC based model.

173bp deletion induces a frame shift in PEAR1 ablating cell surface expression.

PEAR1 KO iPSCs were created using two gRNAs, gRNA-P1E6a and gRNA-P1E6b, to induce a 173 bp deletion leading to a premature termination codon before the transmembrane domain (Figure 3-2). After FACS enrichment for Cas9_GFP+ iPSCs cells were seeded at low density for isolation of individual colonies of which 6/12 screened contained the deletion a remarkable improvement over unsorted cells. After clonal isolation through two rounds of serial dilution, BC1 *PEAR1* KO clones were isolated and a single clone was used for functional studies. Bi-allelic deletion was suspected due to the absence of a wild type size band (Figure 3.2) but further confirmed by the presence of

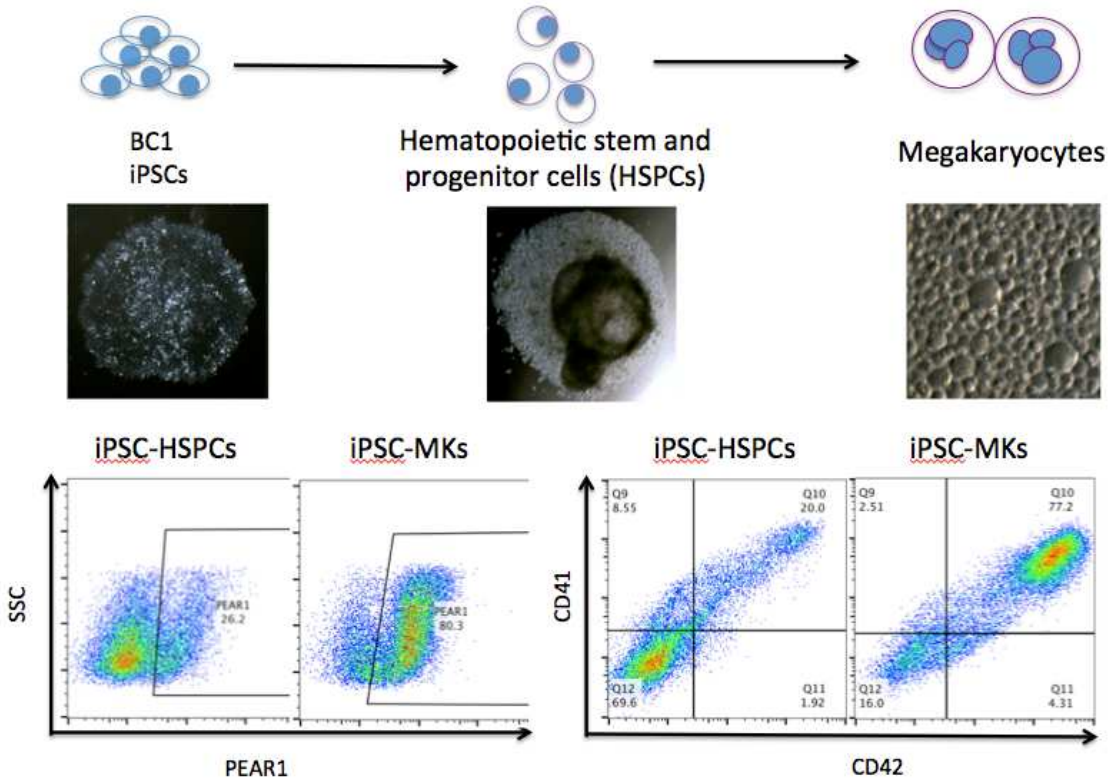
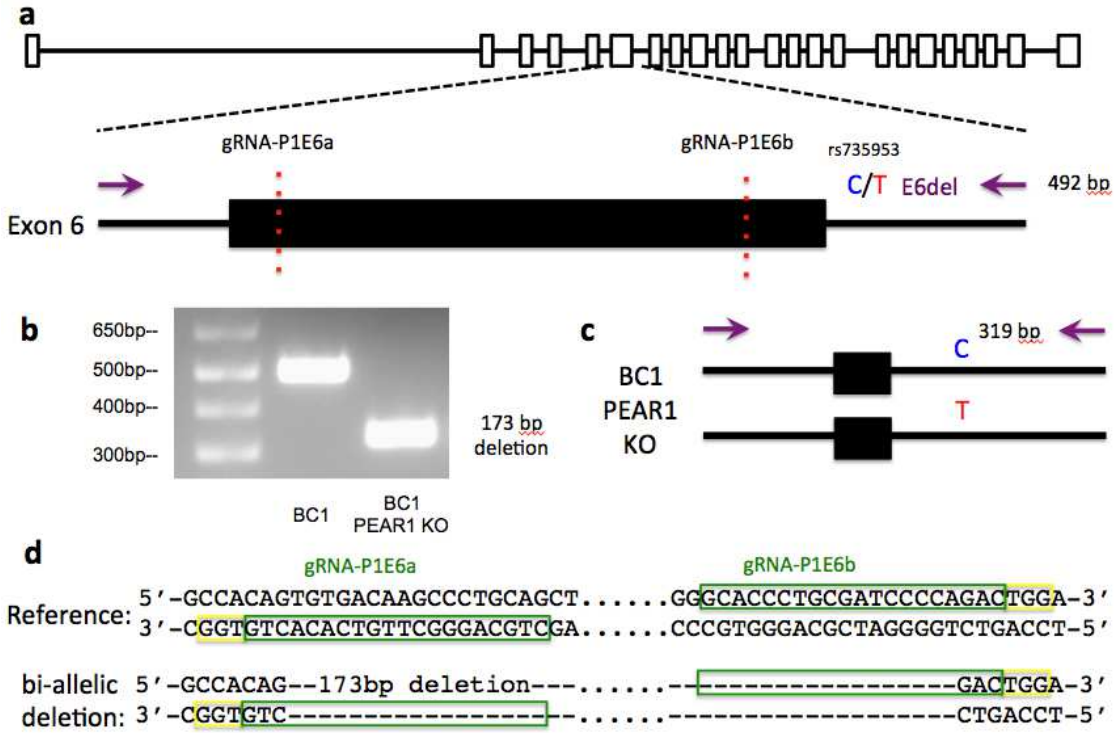


Figure 3-1 An iPSC-based model to generate PEAR1 expressing megakaryocytes. BC1 iPSCs were used for the spin-EB mesoderm differentiation protocol to produce hematopoietic stem and progenitor cells (HSPCs) after 12 days of culture. These HSPCs were found to express PEAR1 at ~25% as detected by FACS and were ~75% double positive for CD41 and CD42, lineage markers for megakaryocytes and their precursors. By day 5 of megakaryocyte differentiation the cells typically become quite large after multiple rounds of DNA replication without cell division.

both alleles of a SNV, rs735953 (T/C), within the sequenced products. By bacterial transformation and colony picking each allele could be sequenced individually revealing the same deletion with breakpoints at the third nucleotide 5' of the PAM sequence for each gRNA present on both the T and the C containing alleles. *PEAR1* KO iPSCs maintained their capacity for hematopoietic differentiation as observed from the release of suspension cells from the embryoid body (EB) and expression of $CD34^{+}CD45^{+}$ double positive cells but had no detectable PEAR1 protein expression at any time point (Figure 3-3). To evaluate mRNA expression of PEAR1 during differentiation qRT-PCR was used for PEAR1 that revealed a 50% reduction in mRNA levels at the iPSC-HSPC stage while a 20% reduction was observed at the iPSC-MK final time point. This indicates that non-sense mediated decay (NMD) or some other mechanism is reducing mRNA levels of the truncated product with a premature termination codon before the final exon.

Figure 3-2 Homozygous PEAR1 KO achieved in BC1 iPSCs. (a) PEAR1 is depicted with exons as boxes and introns as lines with a zoom in on exon 6 for targeted disruption with PCR screening primers shown in purple to generate a 492bp band. Two gRNAs indicated by red dotted lines are spaced 173bp apart within exon 6 (P1E6a and P1E6b). (b) Gel results for E6del PCR showing BC1 iPSCs with the wt 492bp band and BC1 PEAR1 KO a single band the size of anticipated deletion indicating a homozygous deletion. (c) Expected deletion product size after cleavage at both gRNAs yielding a 319 bp band. The unrelated SNP, rs735953 is used to screen the mutation status of each allele after clonal sequencing. (d) The reference sequence is displayed with gRNA sequences in green boxes followed by their protospacer adjacent motif (PAM) in yellow. Both alleles in BC1 PEAR1 KO had the same 173bp deletion starting at the third nucleotide upstream of the PAM from each gRNA.

Figure 3-2 Homozygous PEAR1 KO achieved in BC1 iPSCs.



PEAR1 KO accelerates megakaryocyte lineage commitment and cell growth during iPSC-based differentiation.

During hematopoietic differentiation *PEAR1* KO iPSCs displayed an accelerated emergence of the megakaryocyte lineage as detected by a high expressing population of $CD41^{+}CD42^{+}$ double positive cells when compared to parental iPSCs (Figure 3-5). Hematopoietic lineage markers, CD34 and CD45, were not significantly altered indicating successful blood differentiation was not disrupted in either population. A cell growth phenotype was also observed with a 2.5 fold increased total number of suspension cells harvested from *PEAR1* KO iPSCs when compared with parental cells indicating that PEAR1 acts as a negative regulator of growth during hematopoiesis (Figure 3-4). It was

also observed that release of hematopoietic suspension cells from the embryoid body occurred ~1 day earlier in PEAR1 KO iPSCs when compared to BC1 parental cells. After 5 days of growth under conditions to promote megakaryocyte differentiation BC1 PEAR1 KO iPSCs continued to have a cell growth advantage when compared with parental cells but both populations maxed at ~65% of cells expressing $CD41^{+}CD42^{+}$ double positive cells in addition to the characteristic cell enlargement indicating a successful maturation of megakaryocytes (Figure 3-5).

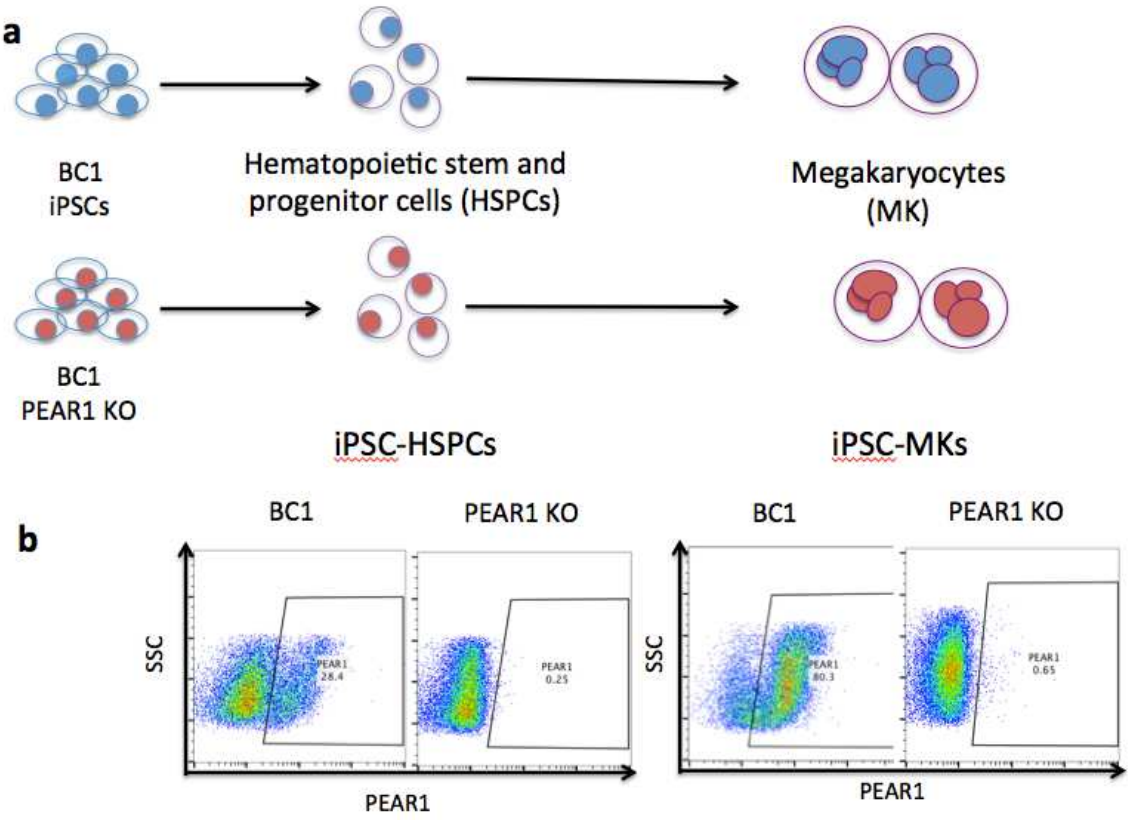


Figure 3-3 BC1 PEAR1 KO ablated PEAR1 protein expression. (**a**) A representation of spin-EB hematopoietic differentiation comparing BC1 iPSCs to BC1 PEAR1 KO iPSCs. Cells are analyzed by isolating suspension cells from the bulk EB at day 12 containing the hematopoietic stem and progenitor (HSPC) which are further differentiated and megakaryocyte stages. (**b**) FACS analysis of HSPCs and MKs display PEAR1 expression on the x-axis by SSC and show the PEAR1 KO iPSC ablated PEAR1 surface expression compared to normal expression from BC1 iPSCs.

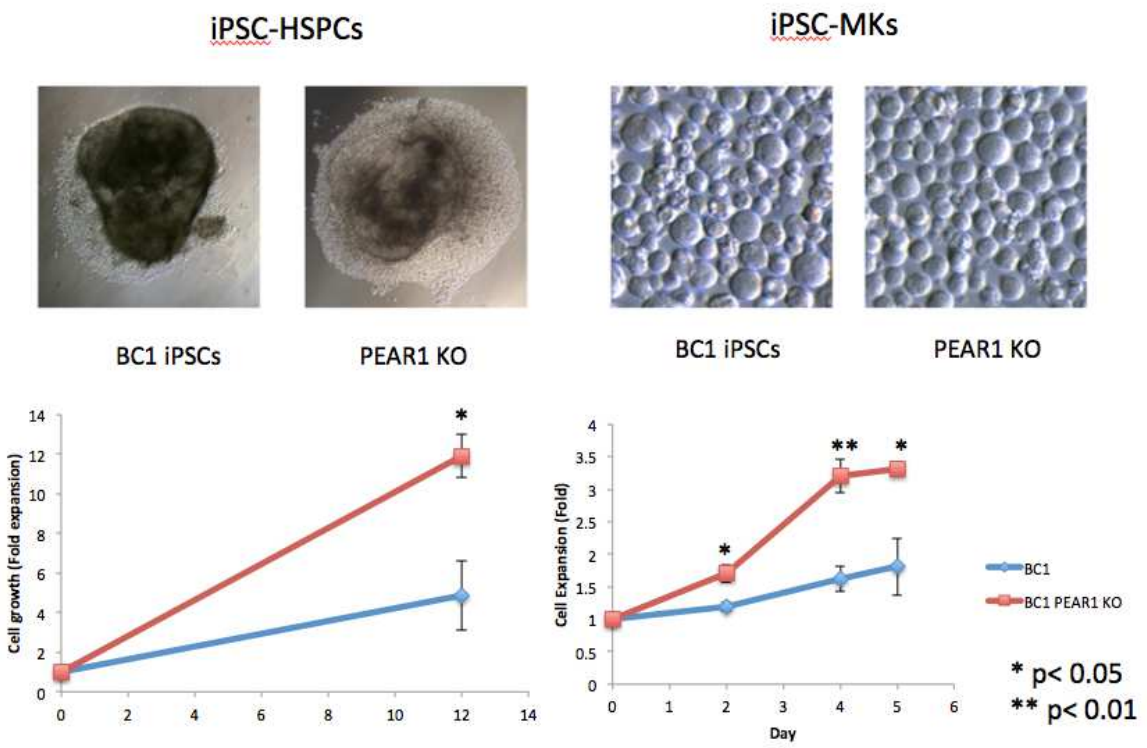


Figure 3-4 PEAR1 KO increased hematopoietic growth. There was a clear growth advantage from the PEAR1 KO when compared to BC1 iPSCs from which they were derived as seen in the early emergence of suspension cells containing the hematopoietic progenitor. A graph of the cell number harvested on day 12 by filtering the bulk EB to only count and analyze cells less than 70 μm used for FACS and continued megakaryocyte differentiation. Data represents an average of three biological replicates of indepenedent differentiation experiments, error bar represents SEM. After 5 days of growth in megakaryocyte inducing media large cells could be observed under the microscope in both BC1 and PEAR1 KO iPSC derived megakaryocytes. Cells were counted at day 2, and 4 when changing media and on day 5 when they were analyzed by FACS. PEAR1 KO had a growth advantage at every time point when compared to BC1 iPSCs.

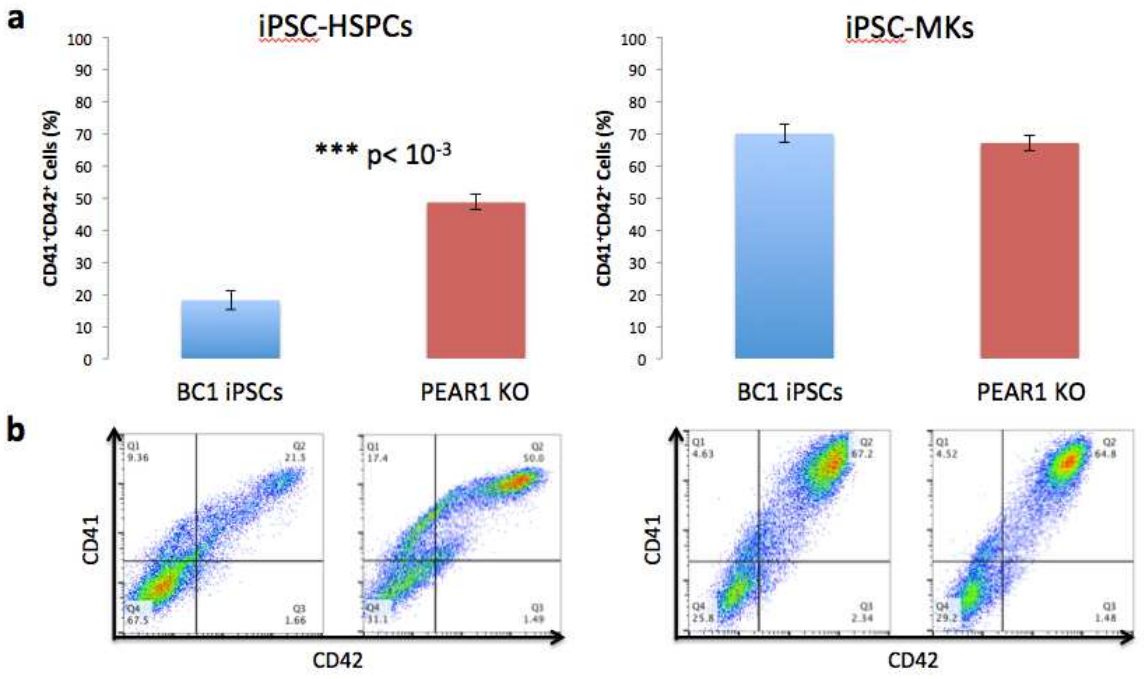


Figure 3-5 PEAR1 KO accelerated megakaryocyte lineage commitment. (**a**) iPSC derived HSPCs and MKs from either BC1 parental iPSCs or PEAR1 KO were screened by FACS and the CD41+CD42+ double positive cells were plotted. Data represents an average of three biological replicates of independent differentiation experiments, error bar represents SEM. (b) FACS plot displaying CD41 on the Y-axis and CD42 on the X-axis. A clear double positive population is observed in PEAR1 KO that is not seen in BC1 until megakaryocyte maturation.

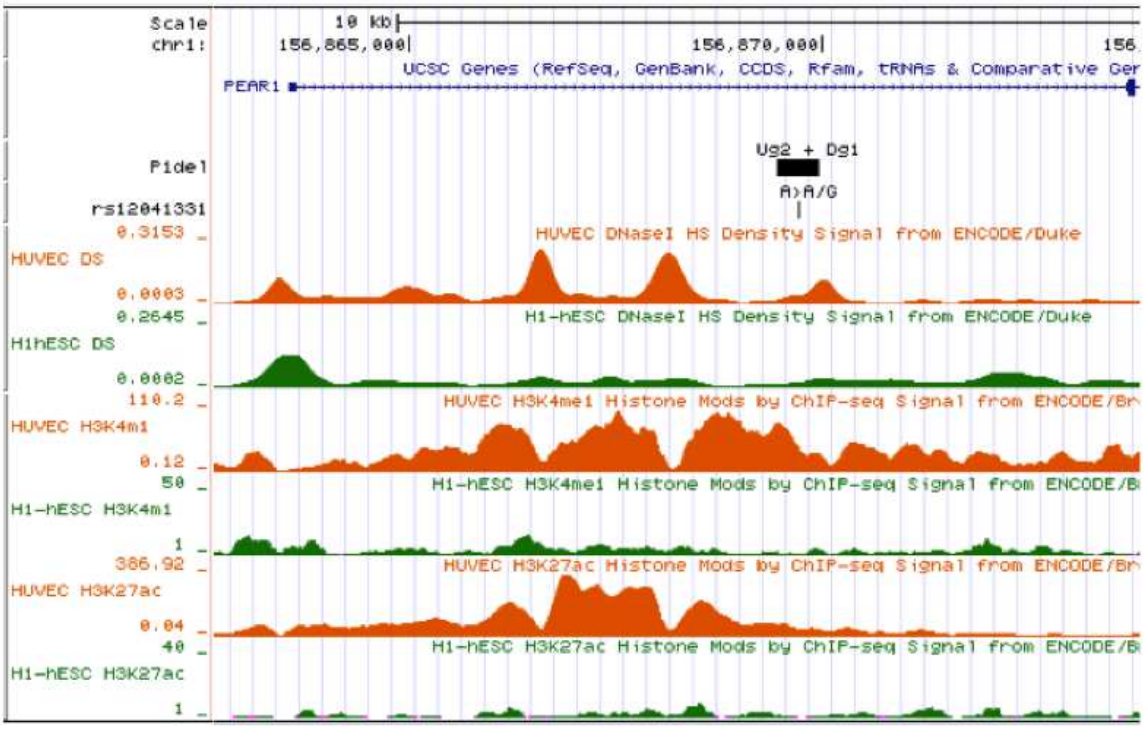


Figure 3-6 Epigenetic landscape of PEAR1 around rs12041331. Data viewed in UCSC genome browser covering the PEAR1 intron 1 shown on the top track. P1del represents the deletion size of gRNAs covering rs12041331 marked on the following track. Orange tracks below represent human umbilical vein endothelial cells (HUVEC) followed by H1 human embryonic stem cells (hESC). DNase hypersensitivity tracks from ENCODE/Duke, along with histone markers including H3K4m1 and H3K27ac.

BC1 cisPEAR1 KO increased cell proliferation during hematopoietic differentiation.

After confirming a functional role for PEAR1 in iPSC derived megakaryopoiesis the regulation of this gene became the target of further study. Genetic variation within the first intron of *PEAR1* showed cis-regulatory potential by the luciferase assay but further characterization of this element in its endogenous genomic context was necessary to elucidate its role in regulation during megakaryopoiesis. To address this BC1 cis*PEAR1* KO was generated using two gRNAs to cause a 251 bp single allele deletion of region surrounding rs12041331, leaving the other allele unmodified by Cas9 (Figure 3-7). Both alleles were isolated by TOPO cloning and sequenced to determine the exact breakpoint and confirm that no mutations occurred on the wild type allele by either gRNA. When BC1 cis*PEAR1* KO was compared to parental BC1 iPSCs, there was a growth advantage during hematopoiesis and megakaryopoiesis (Figure 3-8). The megakaryocyte progenitor population marked by $CD41^{+}CD42^{+}$ double positive cells was slightly but not significantly elevated in iPSC-HSPCs in cisPEAR1 (Figure 3-9) unlike the full PEAR1 KO experiments. Additionally when PEAR1 protein was analyzed by FACS the percentage of cells with cell surface protein was not significantly reduced (Figure 3-10).

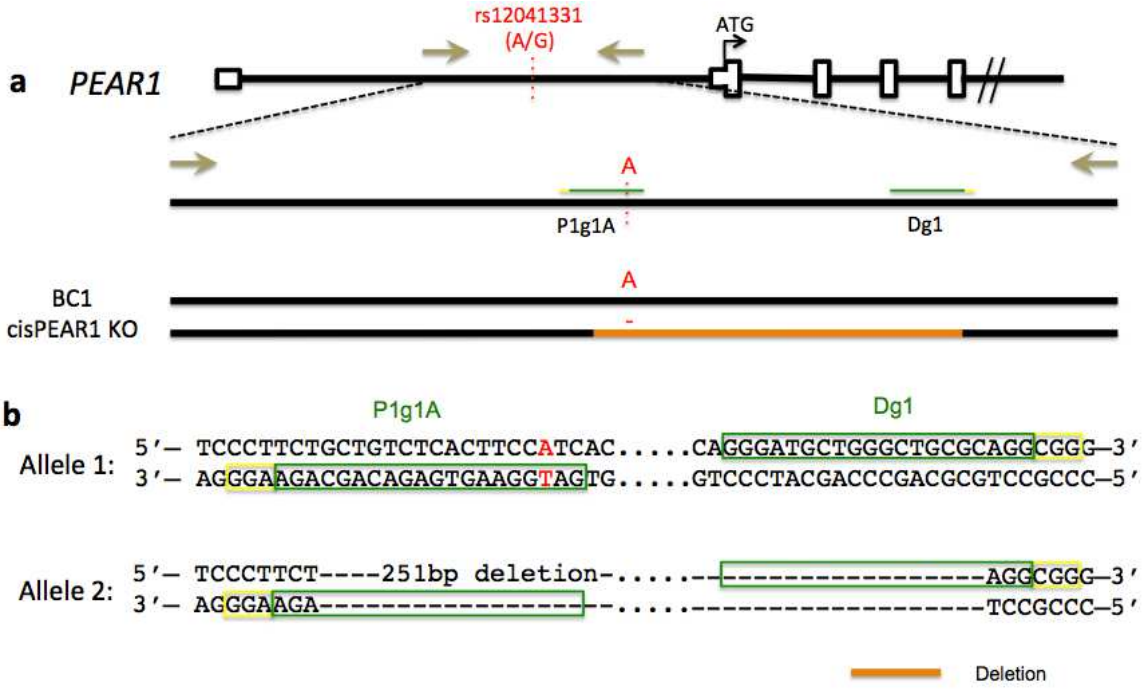


Figure 3-7 Generating cisPEAR1 KO: a deletion of rs12041331 within a putative enhancer. (**a**) The PEAR1 gene is displayed with rs12041331 (A/G) within the first intron expanded for the PCR product used for screening targeted clones. gRNAs (P1g1A and Dg1) are shown as green lines with yellow caps for the PAM sequences. BC1 cisPEAR1 KO is shown below with a wild-type allele containing rs12041331 that is not mutated and the other allele containing a deletion covering this SNP. (**b**) Independent sequencing of each allele showed a 251 bp deletion that started at the third nucleotide upstream of the PAM for each gRNA. The full length allele was shown to harbor no indels at either gRNA site or within the PCR product.

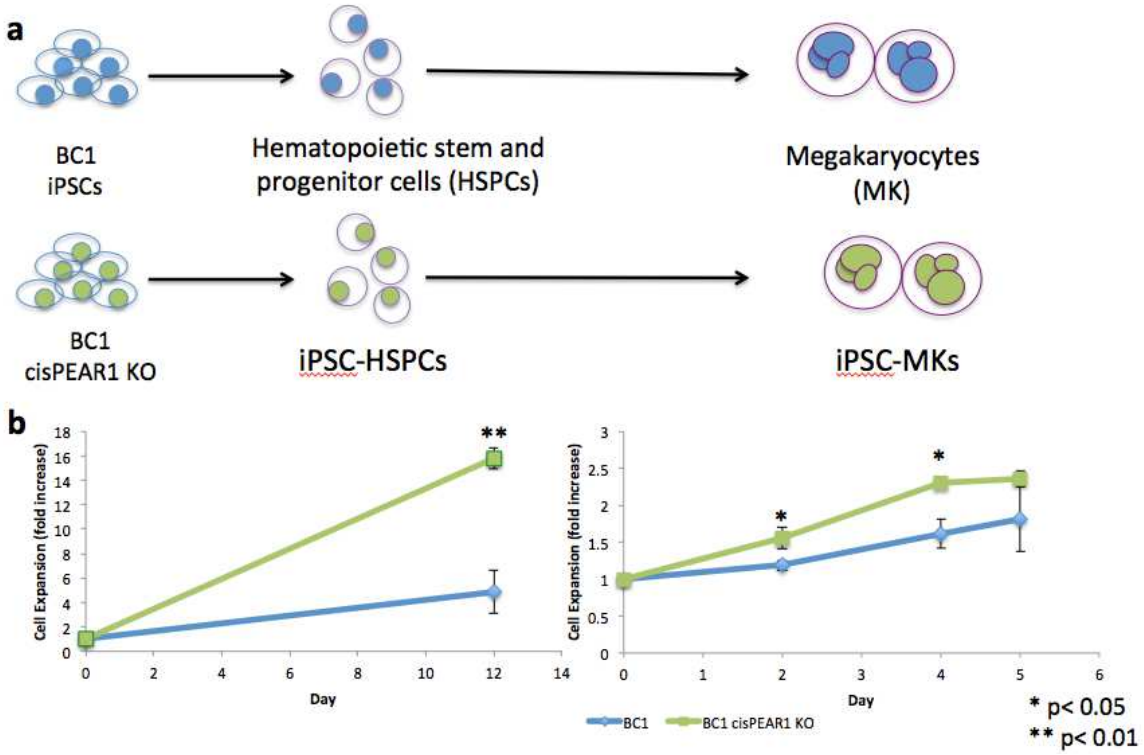


Figure 3-8 cisPEAR1 KO showed increased cell proliferation during hematopoietic differentiation. (**a**) Schematic of spin-EB hematopoietic differentiation followed by megakaryocyte maturation comparing the cisPEAR1 KO clone and the BC1 iPSCs that it was derived from (**b**) Cell growth determined by counting suspension (countess) cells after EB day 12 and counting total cell number during megakaryocyte differentiation. The robust growth phenotype was largely observed in the generation of HSPCs but only a slight increase was observed at two of the three megakaryocyte time points.

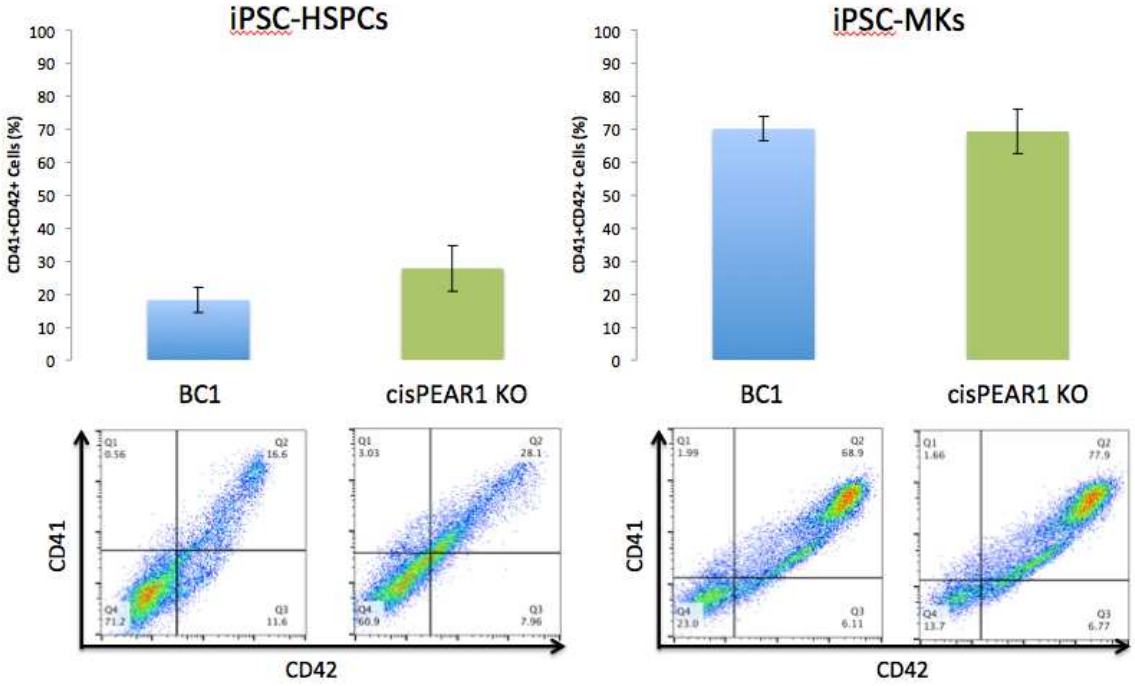


Figure 3-9 Expression of megakaryocyte markers were not significantly elevated in cisPEAR1 KO. (**a**) Percentage of CD41+CD42+ double positive cells as determined by FACS at the HSPC stage and megakaryocyte stage comparing BC1 iPSCs to cisPEAR1 KO. Data represents the mean of three biological replicates with errors bars shown as SEM. (**b**) FACS plot with CD41 on the Y-axis and CD42 and on X-axis. While the overall percentage was slightly but not significantly elevated the plot shows a transitioning population in the cisPEAR1 KO that is not present in BC1 iPSCs.

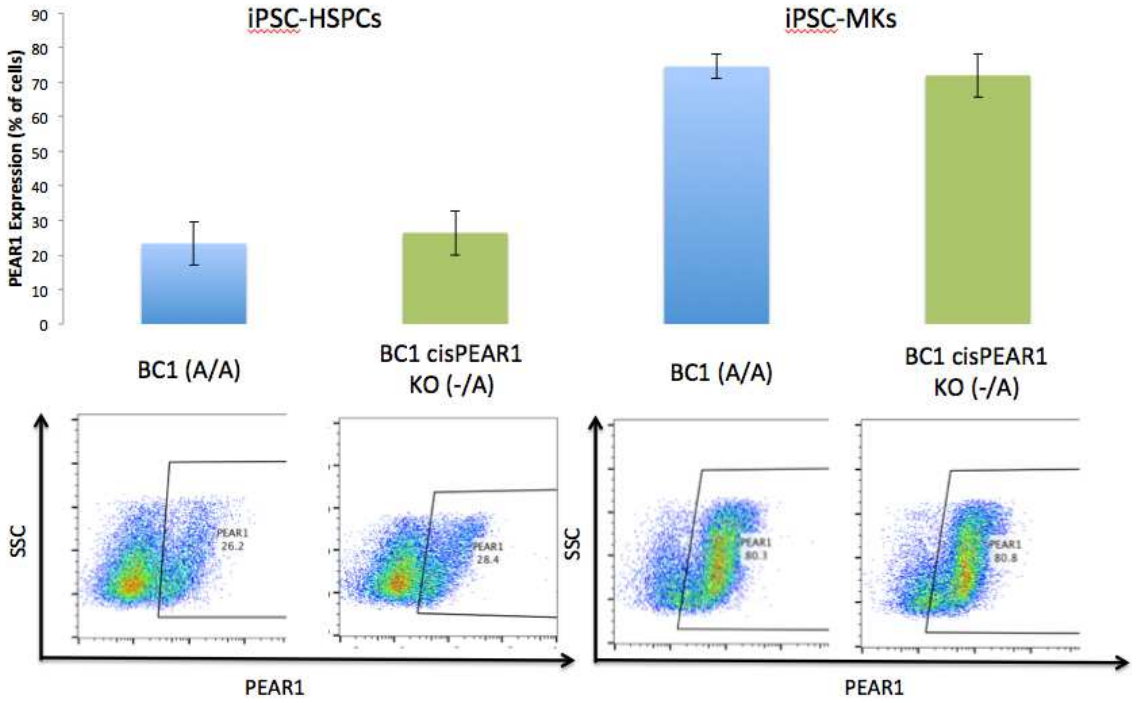


Figure 3-10 cisPEAR1 KO did not affect PEAR1 surface expression. (**a**) Percentage of PEAR1 expression cells as measure by FACS, the mean of 3 experiments with error bars plotted as SEM. (**b**) FACS plot measuring SSC on the Y-axis and PEAR1 on the X-axis. When compared to parental BC1 iPSCs, cisPEAR1 KO did not affect PEAR1 cell surface protein expression.

Deletion of rs12041331 decreased mRNA expression of *PEAR1* on the same allele as the deletion, indicating that it resides within a cis-regulatory element.

When comparing cisPEAR1 KO iPSCs to the BC1 iPSCs that they were derived from *PEAR1* mRNA expression was reduced by ~50% during hematopoietic differentiation and ~30% during megakaryocyte differentiation over three biological

replicates as analyzed by qRT-PCR confirming its regulatory effects during hematopoiesis and megakaryopoiesis (Figure 3-11). Although a decrease in mRNA expression was found in all hematopoietic derived cells, endothelial cells are also of interest as they are known to express PEAR1 and are found in the bulk EB before the hematopoietic suspension cells are released. When bulk EB cells were analyzed after removing all free cells with a 70 μm filter PEAR expression was quantified using qRT-PCR and found to express PEAR1 ~50 fold greater than HSPCs (Figure 3-12). Interestingly there was no significant difference in PEAR1 expression between BC1 and cisPEAR1 KO in the bulk EB cells indicating that this enhancer is cell type specific and does not influence gene expression in this higher expressing mix of cell types that make up the EB.

To determine if the reduction in mRNA was occurring on the same allele as the enhancer deletion a naturally occurring unrelated SNP, rs77235035 (C/A), within exon 8 was utilized to monitor the ratio of PEAR1 expression from each allele. The linkage of the intronic SNP of interest rs12041331 was linked to exon 8 by long range PCR which was TOPO cloned and sequenced to isolate each allele. Total mRNA extracted from BC1 iPSCs during differentiation was first converted to cDNA then PCR amplified around spliced exon 8 and sequenced resulting in 79.5% of reads contained the A allele indicating a baseline bias that influences PEAR1 expression from a specific allele in iPSC-HSPCs and Megakaryocytes (Figure 3-13). As BC1 is homozygous AA for rs12041331 there is likely other regulatory elements that are influencing PEAR1 expression in an allele specific manor that are not identified or covered in this study.

When BC1 cis*PEAR1* KO mRNA was analyzed from iPSC-derived HSPCs the A allele was only present in 32.0% of the reads sequenced indicating the reduction of total mRNA is primarily occurring on the same allele as the intronic deletion suggesting a cis-regulatory mechanism of action for this element at this stage of development. Upon analyzing iPSC-derived megakaryocytes from BC1 cis*PEAR1* KO the A allele was sequenced in 28.7% of reads, a significant reduction from BC1 parental cells.

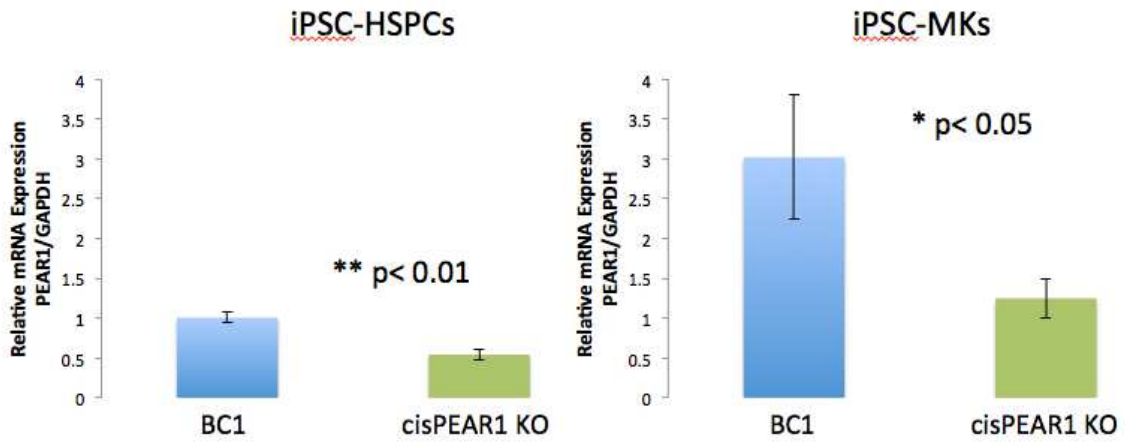


Figure 3-11 PEAR1 total mRNA levels were reduced during differentiation in cisPEAR1 KO. Cell pellets were harvested from both BC1 and cisPEAR1 KO at the HSPC and Megakaryocyte time points for RNA extraction, and cDNA conversion for quantitative RT-PCR. Expression was normalized to GAPDH for each sample and normalized to BC1 HSPCs to a level of one for comparison across all samples. A 50% reduction was observed at both the HSPC and the MK time points but more variation was observed in the megakaryocytes. Data displayed is the mean of three biological replicates with the SEM plotted for error bars.

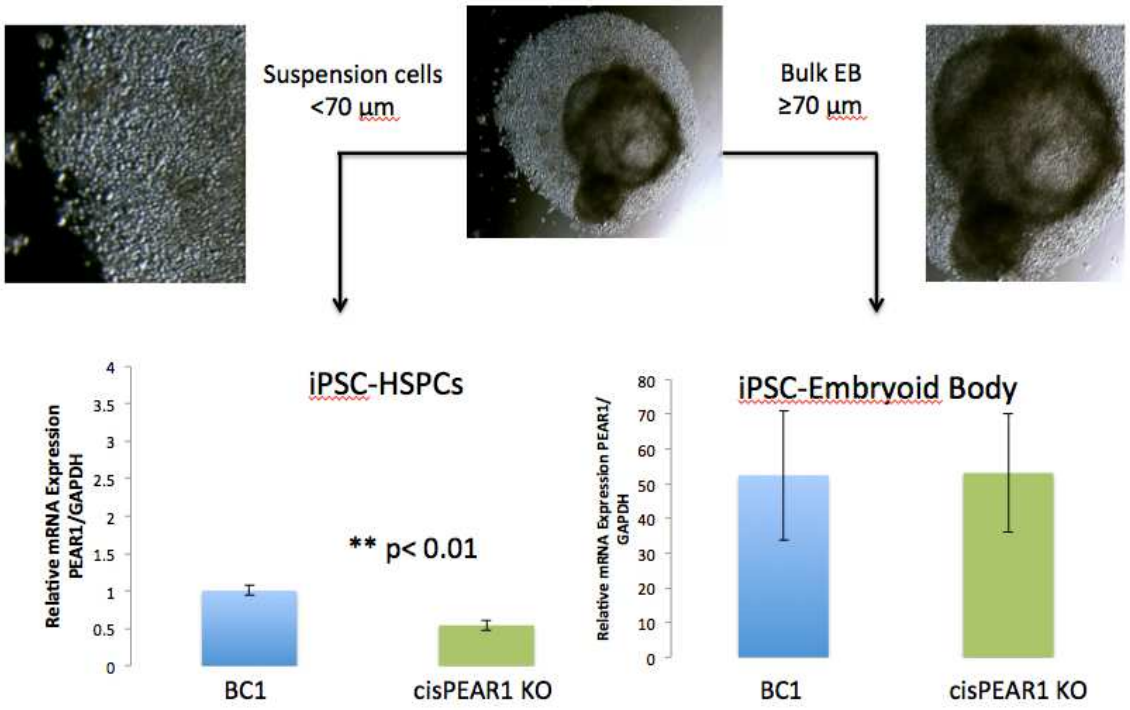


Figure 3-12 PEAR1 mRNA levels were not reduced in bulk embryoid body cells derived from cisPEAR1 KO. At day 12 of spin-EB hematopoietic differentiation the suspension cells were harvested using a 70 µm filter while the cellular material not filtered was collected for bulk EB analysis. RT-PCR was conducted as previously described to reveal an overall higher level of expression of PEAR1 in the bulk EB cells when compared to hematopoietic derived cells. Additionally the cisPEAR1 KO did not show reduced PEAR1 expression in the bulk cells but only the HSPCs and MKs indicating a cell type specific effect of this enhancer.

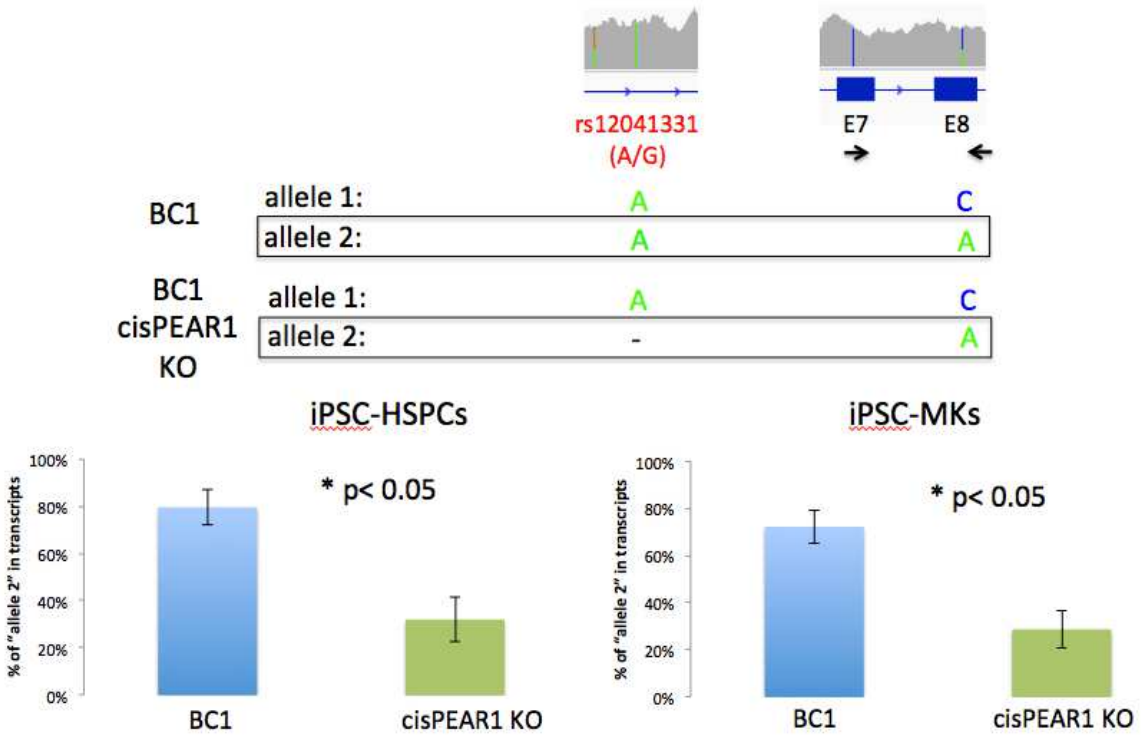


Figure 3-13 cisPEAR1 KO decreased PEAR1 mRNA on the same allele as the deletion. (**a**) PEAR1 genetic variation is displayed from the IGV genome browser showing BC1 genetic variation from the reference. BC1 is homozygous A for rs12041331, key SNP for this study but an unrelated SNP, rs77235035, within exon 8 was used to monitor allelic expression of PEAR1. Using long range PCR both BC1 iPSCs and cisPEAR1 KO were sequenced to determine the linkage group for each allele. "Allele 2" for BC1 cisPEAR1 KO is monitored by the A allele of Exon 8. (**b**) Plots displaying the percentage of "allele 2" (A) reads from mRNA divided by the total reads for BC1 and cisPEAR1 KO from three independent transfections, error bars are SEM.

AIM 3 DISCUSSION

The proper biosynthesis and functionality of platelets is integral to hemostasis, wound recovery and is involved in many complex diseases such as atherosclerosis. GWAS have identified a plethora of variants for a wide range of human phenotypes and diseases but proper identification of the causal variant and functional follow up studies are often lacking or missing all together. While increasing the cohort size will reveal variants of smaller effect, multigenic or gene environment interactions may be impossible to discern using this purely statistical method. With the rapid increases in precise genome editing it is now feasible to alter these variants of interest and evaluate their effects in isolation or under certain controlled conditions such as differentiation or chemical induction. These results demonstrate the feasibility for CRISPR/Cas9 based genome editing techniques to either disrupt a protein coding gene or putative cis-regulatory element associated with biological phenotype to discern its cellular and molecular effects.

This study first showed that an iPSC based model of hematopoiesis and megakaryopoiesis mirrors biological development in the proper timing and expression of PEAR1 and the megakaryocyte lineage markers CD41 and CD42; a requirement to investigate PEAR1 function in this *in vitro* system. This differentiation model provided an ideal environment to KO PEAR1 and its putative enhancer containing the SNP rs12041331 to interrogate their effects on megakaryopoiesis or PEAR1 mRNA and protein expression. The first major results confirm *PEAR1* as a non-essential negative regulator of megakaryopoiesis as a cell growth advantage was observed in BC1 *PEAR1* KO when compared with otherwise isogenic parental controls but production of mature

megakaryocytes was not disrupted. This was not surprising as *PEAR1* was shown to signal through the PI3K/AKT pathway, which is known to regulate cell proliferation and was previously reported to negatively influence cell growth in human $CD34^+$ peripheral blood cells during megakaryocyte differentiation (Kauskot et al. 2013). While these observations are in concordance with the Kasuskot study, an iPSC based approach also undergoes mesoderm lineage commitment to produce the HSPCs necessary for megakaryocyte differentiation, an earlier stage in development than the previous studies using peripheral blood derived $CD34^+$ cells. A cell proliferation advantage was also observed during this early hematopoietic differentiation indicating PEAR1 influences commitment and or cell proliferation at an earlier stage in development than previously reported. While previous investigation into the function PEAR1 and its regulation relied on synthetic reporter constructs or siRNA knockdown approaches this study used CRISPR/Cas9 guided deletion determine the role that this gene and its putative enhancer play in the regulation megakaryopoiesis.

After probing the role of the PEAR1 gene by coding sequence KO the next objective of this aim was to follow-up on the putative enhancer within its first intron containing the SNP rs12041331. Previous studies provided several lines of evidence that this region was acting as a regulatory of gene expression from peripheral blood analysis as well as luciferase assay, but has not been replicated since (Faraday 2011). By deleting 251 bp containing rs12041331 an overall reduction in mRNA levels was observed in both iPSC derived HSPCs and megakaryocytes confirming the previous observations but in the endogenous chromatin context of iPSCs differentiated to the relevant cell type. In the

bulk EB total mRNA levels were ~50 x higher which corroborates previous reports that endothelial cells express PEAR1 at orders of magnitude higher levels than hematopoietic derived cells (Nanda et al. 2005) but interestingly the levels of PEAR1 expression in the cisPEAR1 KO clone was not reduced in this cell type but only in the hematopoietic lineages. This reveals the regulatory potential of this region on the PEAR1 gene within which it resides in a cell type specific manor but from totals mRNA levels a cis-acting mechanism of regulation cannot be distinguished from other indirect mechanisms of gene regulation that might act on both alleles equally. By deep sequencing mRNA from HSPCs and MKs derived from BC1 iPSCs it was determined that the exon 8 variant A for rs77235035, allele 2, was expressed four times greater than the C containing allele 1 (Figure 3-13). This indicates a pre-existing bias in allelic expression and as BC1 is homozygous for rs12041331, there is likely other regulatory regions or an imprinting effect that is influencing PEAR1 expression in an allele specific manor independent of any potential enhancer effects identified by deletion of rs12041331. When cisPEAR1 KO was compared to BC1 iPSCs there was a reduction in expression found on the same allele as the deletion of rs12041331 bring relative expression down to ~30% for allele 2 when compared to ~75% for BC1. The reduction in total mRNA levels along with a change in the ratio of PEAR1 allelic expression is supportive of a cis-acting enhancer that acts in hematopoietic derived cells including megakaryocytes.

The regulation of hematopoiesis and proper production of a full complement of blood derived cell lineages is a critical component of hemostasis, and general well being. Overproduction or lack of certain blood lineages has been reported in both Mendelian

disease as well as in cancer and through the study of rare forms of hematologic disease many key genes such as RUNX1, GATA1 or DBA in which mutations have been reported to cause highly penetrant forms of disease but variants of smaller impact have been difficult to identify and study. While initially PEAR1 was reported in the molecular mechanism of platelet aggregation; its more recent implication in regulating hematopoiesis and megakaryopoiesis made it an attractive target for an iPSC based genome editing protocol to further interrogate its function confirming its previously reported role in regulating growth and lineage commitment during megakaryopoiesis while also revealing an allele specific mechanism of gene regulation for an enhancer within its first intron containing the SNP rs12041331.

Bibliography

1. Gurdon JB, Elsdale TR, Fischberg M. (1958). Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. *Nature* **182**:64-65.
2. Takahashi K, Yamanaka S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**:663-676.
3. Zou, J, Maeder, ML, Mali, P, Pruett-Miller, SM, Thibodeau-Beganny, S, Chou, BK et al. (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5**: 97–110.
4. Hockemeyer, D, Soldner, F, Beard, C, Gao, Q, Mitalipova, M, DeKelver, RC et al. (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* **27**: 851–857.
5. Hockemeyer, D, Wang, H, Kiani, S, Lai, CS, Gao, Q, Cassady, JP et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* **29**: 731–734.
6. Mali, P, Yang, L, Esvelt, KM, Aach, J, Guell, M, DiCarlo, JE et al. (2013). RNA-guided human genome engineering via Cas9. *Science* **339**: 823–826.
7. Cong, L, Ran, FA, Cox, D, Lin, S, Barretto, R, Habib, N et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819–823.
8. Damian, M and Porteus, MH (2013). A crisper look at genome editing: RNA-guided genome modification. *Mol Ther* **21**: 720–722.
9. Jinek, M, Chylinski, K, Fonfara, I, Hauer, M, Doudna, JA and Charpentier, E (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816–821.
10. Levine, RL and Gilliland, DG (2008). Myeloproliferative disorders. *Blood* **112**: 2190–2198.
11. Carrell, RW and Lomas, DA (2002). Alpha1-antitrypsin deficiency–a model for conformational diseases. *N Engl J Med* **346**: 45–53.
12. Smith, JR, Maguire, S, Davis, LA, Alexander, M, Yang, F, Chandran, S et al. (2008). Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. *Stem Cells* **26**: 496–504.
13. Lombardo, A, Cesana, D, Genovese, P, Di Stefano, B, Provasi, E, Colombo, DF et al. (2011). Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods* **8**: 861–869.
14. Zou, J, Sweeney, CL, Chou, BK, Choi, U, Pan, J, Wang, H et al. (2011). Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. *Blood* **117**: 5561–5572.
15. Zou, C, Chou, BK, Dowey, SN, Tsang, K, Huang, X, Liu, CF et al. (2012). Efficient derivation and genetic modifications of human pluripotent stem cells on engineered human feeder cell lines. *Stem Cells Dev* **21**: 2298–2311.
16. Hsu, PD, Scott, DA, Weinstein, JA, Ran, FA, Konermann, S, Agarwala, V et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**: 827–832.
17. Fu, Y, Foden, JA, Khayter, C, Maeder, ML, Reyon, D, Joung, JK et al. (2013). High- frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**: 822–826.
18. Niu, Y, Shen, B, Cui, Y, Chen, Y, Wang, J, Wang, L et al. (2014). Generation of gene- modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* **156**: 836–843.
19. Wu, Y, Liang, D, Wang, Y, Bai, M, Tang, W, Bao, S et al. (2013). Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell* **13**: 659–662.
20. Schwank, G, Koo, BK, Sasselli, V, Dekkers, JF, Heo, I, Demircan, T et al. (2013). Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13**: 653–658.
21. Wu, X, Scott, DA, Kriz, AJ, Chiu, AC, Hsu, PD, Dadon, DB et al. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* **32**: 670–676.
22. Yang, H, Wang, H, Shivalila, CS, Cheng, AW, Shi, L and Jaenisch, R (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**: 1370–1379.

23. Yang, L, Guell, M, Byrne, S, Yang, JL, De Los Angeles, A, Mali, P et al. (2013). Optimization of scarless human stem cell genome editing. *Nucleic Acids Res* **41**: 9049–9061.
24. Yin, H, Xue, W, Chen, S, Bogorad, RL, Benedetti, E, Grompe, M et al. (2014). Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat Biotechnol* **32**: 551–553.
25. Smith, C, Gore, A, Yan, W, Abalde-Atristain, L, Li, Z, He, C et al. (2014). Whole- genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* **15**: 12–13.
26. Veres, A, Gosis, BS, Ding, Q, Collins, R, Ragavendran, A, Brand, H et al. (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* **15**: 27–30.
27. Suzuki, K, Yu, C, Qu, J, Li, M, Yao, X, Yuan, T et al. (2014). Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. *Cell Stem Cell* **15**: 31–36.
28. Chou, BK, Mali, P, Huang, X, Ye, Z, Dowey, SN, Resar, LM et al. (2011). Efficient human iPS cell derivation by a non-integrating plasmid from blood cells with unique epigenetic and gene expression signatures. *Cell Res* **21**: 518–529.
29. Cheng, L, Hansen, NF, Zhao, L, Du, Y, Zou, C, Donovan, FX et al.; NISC Comparative Sequencing Program. (2012). Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. *Cell Stem Cell* **10**: 337–344.
30. Choi, SM, Kim, Y, Shim, JS, Park, JT, Wang, RH, Leach, SD et al. (2013). Efficient drug screening and gene correction for treating liver disease using patient-specific stem cells. *Hepatology* **57**: 2458–2468.
31. Yan, W, Smith, C and Cheng, L (2013). Expanded activity of dimer nucleases by combining ZFN and TALEN for genome editing. *Sci Rep* **3**: 2376.
32. Porteus, MH (2006). Mammalian gene targeting with designed zinc finger nucleases. *Mol Ther* **13**: 438–446.
33. Yusa, K, Rashid, ST, Strick-Marchand, H, Varela, I, Liu, PQ, Paschon, DE et al. (2011). Targeted gene correction of a1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478**: 391–394.
34. Ye, Z, Zhan, H, Mali, P, Dowey, S, Williams, DM, Jang, YY et al. (2009). Human- induced pluripotent stem cells from blood cells of healthy donors and patients with acquired blood disorders. *Blood* **114**: 5473–5480.
35. Choi, SM, Liu, H, Chaudhari, P, Kim, Y, Cheng, L, Feng, J et al. (2011). Reprogramming of EBV-immortalized B-lymphocyte cell lines into induced pluripotent stem cells. *Blood* **118**: 1801–1805.
36. Ye, Z, Liu, CF, Lanikova, L, Dowey, SN, He, C, Huang, X et al. (2014). Differential sensitivity to JAK inhibitory drugs by isogenic human erythroblasts and hematopoietic progenitors generated from patient-specific induced pluripotent stem cells. *Stem Cells* **32**: 269–278.
37. Ding, Q, Regan, SN, Xia, Y, Oostrom, LA, Cowan, CA and Musunuru, K (2013). Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* **12**: 393–394.
38. Li, T, Huang, S, Jiang, W, Wright, D, Spalding, M Weeks, D and Yang, B. (2011). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res* **39**: 359–372.
39. Cermak, T, Doyle, E, Christian, M, Wang, L, Zhang, Y, Schmidt, C, Baller, J, Somia, N, Bogdanove, A, Voytas, D. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**: e82.
40. Djebali, S, Davis, CA, Merkel, A, Dobin, A, Lassmann, T, Mortazavi, A et al. (2012). Landscape of transcription in human cells. *Nature* **489**: 101–108.
41. Soldner, F, Laganière, J, Cheng, AW, Hockemeyer, D, Gao, Q, Alagappan, R et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* **146**: 318–331.
42. Ramirez, CL, Certo, MT, Mussolino, C, Goodwin, MJ, Cradick, TJ, McCaffrey, AP et al. (2012). Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res* **40**: 5560–5568.

43. Ran, FA, Hsu, PD, Lin, CY, Gootenberg, JS, Konermann, S, Trevino, AE et al. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**: 1380–1389.
44. Shen, B, Zhang, W, Zhang, J, Zhou, J, Wang, J, Chen, L et al. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods* **11**: 399–402.
45. Maresca, M, Lin, VG, Guo, N and Yang, Y (2013). Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res* **23**: 539–546.
46. Hendel, A, Kildebeck, EJ, Fine, EJ, Clark, JT, Punjya, N, Sebastiano, V et al. (2014). Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell Rep* **7**: 293–305.
47. Chen, G, Gulbranson, DR, Hou, Z, Bolin, JM, Ruotti, V, Probasco, MD et al. (2011). Chemically defined conditions for human iPSC derivation and culture. *Nat Methods* **8**: 424–429.
48. Wang, Y, Chou, BK, Dowey, S, He, C, Gerecht, S and Cheng, L (2013). Scalable expansion of human induced pluripotent stem cells in the defined xeno-free E8 medium under adherent and suspension culture conditions. *Stem Cell Res* **11**: 1103–1116.
49. Reyon, D, Khayter, C, Regan, MR, Joung, JK and Sander, JD (2012). Engineering designer transcription activator-like effector nucleases (TALENs) by REAL or REAL-Fast assembly. Curr *Protoc Mol Biol* Chapter 12: Unit 12.15.
50. Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**:1262–1278.
51. Nanda N, Bao M, Lin H, Clauser K, Komuves L, Quertermous T, Conley PB, Phillips DR, Hart MJ. (2005). Platelet endothelial aggregation receptor 1 (PEAR1), a novel epidermal growth factor repeat-containing transmembrane receptor, participates in platelet contact-induced activation. *J Biol Chem* **280**:24680-24689.
52. Faraday N, Yanek LR, Yang XP, Mathias R, Herrera-Galeano JE, Suktitipat B, Qayyum R, Johnson AD, Chen MH, Tofler GH, Ruczinski I, Friedman AD, Gylfason A, Thorsteinsdottir U, Bray PF, O'Donnell CJ, Becker DM, Becker LC. (2011). Identification of a specific intronic PEAR1 gene variant associated with greater platelet aggregability and protein expression. *Blood* **118**:3367-3375.
53. Kauskot A, Di Michele M, Loyen S, Freson K, Verhamme P, Hoylaerts MF. (2012). A novel mechanism of sustained platelet αIIbβ3 activation via PEAR1. *Blood* **119**:4056-4065.
54. Lewis JP, Ryan K, O'Connell JR, Horenstein RB, Damcott CM, Gibson Q, Pollin TI, Mitchell BD, Beitelshees AL, Pakzy R, Tanner K, Parsa A, Tantry US, Bliden KP, Post WS, Faraday N, Herzog W, Gong Y, Pepine CJ, Johnson JA, Gurbel PA, Shuldiner AR. (2013). Genetic variation in PEAR1 is associated with platelet aggregation and cardiovascular outcomes. *Circ Cardiovasc Genet* **6**:184-192.
55. Kauskot A, Vandenbriele C, Louwette S, Gijsbers R, Tousseyn T, Freson K, Verhamme P, Hoylaerts MF. (2013). PEAR1 attenuates megakaryopoiesis via control of the PI3K/PTEN pathway. *Blood* **121**:5208-5217.
56. Kim Y, Suktitipat B, Yanek LR, Faraday N, Wilson AF, Becker DM, Becker LC, Mathias RA. (2013). Targeted deep resequencing identifies coding variants in the PEAR1 gene that play a role in platelet aggregation. *PLoS One* **8**:e64179.
57. Würtz M, Nissen PH, Grove EL, Kristensen SD, Hvas AM. (2014) Genetic determinants of on-aspirin platelet reactivity: focus on the influence of PEAR1. *PLoS One* **9**:e111816.
58. Liu Y, Wang Y, Gao Y, Forbes JA, Qayyum R, Becker L, Cheng L, Wang Z. (2015) Efficient generation of megakaryocytes from human induced pluripotent stem cells using food and drug administration-approved pharmacological reagents. *Stem Cells Transl Med* **4**:309-319.

Curriculum Vitae

Cory James Smith

Born in Gainesville, Florida USA ● Cell: 352 538 2649 ● CorySmith1618@gmail.com

EDUCATION

Bachelor of Science in Microbiology and Cell Science August 2005 - December 2008
University of Florida, College of Liberal Arts and Sciences: Gainesville, FL
Summa Cum Laude, Cumulative GPA 3.95

Predoctoral Training Program in Human Genetics August 2010 - Current
Johns Hopkins University, School of Medicine: Baltimore, MD

EXPERIENCE

Graduate Student August 2010 – Current
Human Genetics Program: Johns Hopkins University School of Medicine
Linzhao Cheng, Ph.D., Professor of Medicine, Division of Hematology, Institute of Cell Engineering

Compared the efficiency and specificity of ZFNs, TALEN, and CRISPR/Cas9 in human iPSCs.
TALEN and Cas9 were equivalent in their stimulation of HR while Cas9 had a significant increase in indel disruption through NHEJ and predetermined deletions using two gRNAs.
Targeted sequencing of off-target loci reveals higher specificity of Cas9 in iPSCs compared to 293Ts.
Whole genome sequencing of CRISPR/Cas9 and TALEN targeted iPSC clones identified ~200 SNVs and ~10 indels but none were recurrent or similar to the nuclease binding site indicating high specificity.
Disease Allele Specific targeting in patient derived iPSCs for disruption or correction.
Designed gRNAs targeting either the mutant or the wild-type allele and examined their efficiency in editing the intended target allele or the other allele differing by a single nucleotide alone.
Targeting JAK2 V617F gRNAs were highly specific for a single nucleotide for NHEJ and HR.
Functional interrogation of genetic variation via precise genome editing in human stem cells.
Genetic variation (rs12041331) within *Platelet and Endothelial Aggregation Receptor 1* (*PEAR1*) was previously associated with platelet aggregation and increased cardiovascular event rates.
PEAR1 KO resulted in accelerated megakaryocyte commitment during hematopoietic differentiation.
Single allele deletion of rs12041331 within the first intron reduced *PEAR1* mRNA expression on the same allele as the deletion implying it resides within a cis-regulatory element.

Research Internship, J. Craig Venter Institute: Microbial & Environmental Genomics
Robert Coyne, Ph.D., Staff Scientist Summer 2008

Metagenomic sequencing and analysis of *Ichthyopthyrius multifiliis* and its endosymbiotic bacterial population.
Prepared, ran, analyzed, and assembled Sanger Sequencing reactions, and long range PCR.
A project originally designed to sequence the eukaryote primarily returned bacterial endosymbiont reads.
Designed primers and ran combinatorial PCR to close gaps in the individual bacterial genomes.
Genome closure of several strains of *E. coli* O157:H7 responsible for the spinach outbreak in 2006.
Performed transposon bombing to resolve repetitive regions caused by multiple viral insertions.

Research Assistant, Center for Molecular Microbiology: UF Health Sciences Center
Ann Progulske-Fox, Ph.D., Distinguished Professor and Director 2006 – 2008

The role of oral microbiota in the onset and development of atherosclerosis and coronary artery disease
Performed adherence and cell invasion assays on 5 strains of the anaerobic *Porphyromonas gingivalis.*
De novo sequencing and genome closure of *P. gingivalis* using pyrosequencing.
Signal Transduction: Environmental Stimulus to Changes in Global Transcription in Porphyromonas gingivalis.
Used neural network bioinformatics software to identify genes involved in genetic regulation.
Identified the two-component regulatory systems of *P. gingivalis* unique to virulent strains.

Undergraduate Research, Interdisciplinary Center for Biotechnology Research
Eric Triplett, Ph.D., Professor and Chair, Department of Microbiology and Cell Science
The University of Florida

Whole genome shotgun sequencing of *Enterobacter clocace* P101 using 454 pyrosequencing
E. clocace is a plant endophyte that increases plant growth and improves nitrogen nutrition.
Start to finish process from DNA isolation, library construction, sequencing, assembly and analysis.
Annotation and analysis of the 5.3 Mbp genome of *Enterobacter clocace* P101
Predicted protein-coding regions, tRNA and rRNA genes, promoters, terminators, and operon units.
Used FGENESB to identify regions and Artemis to manually annotate regions of interest.

PUBLICATIONS

- Zou C, Chou BK, Dowey SN, Tsang K, Huang X, Liu CF, Smith C, Yen J, Mali P, Zhang YA, Cheng L, Ye Z (2012) Efficient derivation and genetic modifications of human pluripotent stem cells on engineered human feeder cell lines. *Stem Cells Dev*, **21**:2298-2311.
- Yan W, Smith C, Cheng L (2013) Expanded activity of dimer nucleases by combining ZFN and TALEN for genome editing. *Scientific Reports*, **3**:2376.
- Cai J, Miao X, Li Y, Smith C, Tsang K, Cheng L and Wang QF. (2014) Whole Genome Sequencing Identifies Genetic Variances in Culture-Expanded Human Mesenchymal Stem Cells. *Stem Cell Reports **3**:227-233.*
- Smith C, Gore A, Yan W, Abalde-Atristain L, Li Z, He C, Wang Y, Brodsky RA, Zhang K, Cheng L and Ye Z (2014) Whole genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell*, **15**:12-13.
- Smith C, Abalde-Atristain L, He C, Brodsky BR, Braunstein EM, Chaudhari P, Jang YY, Cheng L and Ye Z (2015) Efficient and allele-specific genome editing of disease loci in human iPSCs. *Molecular Therapy*, **23**:570-577.
- Huang X, Wang Y, Yan W, Smith C, Ye Z, Wang J, Gao Y, Mendelsohn L, Cheng L (2015). Corrected adult β-globin protein production in mature erythrocytes differentiated from patient iPSCs after genome editing of the sickle point mutation. *Stem Cells*, **33**:1470-1479.

PRESENTATION

American Society of Gene and Cell Therapy (ASGCT) 17th Annual Meeting May 2014
"High Efficiency and Specificity of Human Genome Editing by Cas9-gRNA Targeting Disease-Associated Loci in Patient-Specific iPSCs".

ACADEMIC AWARDS

American Society of Gene and Cell Therapy Meritorious Abstract Travel Award	May 2014
University of Florida Six Time President's Honor Roll Recipient (4.0 GPA)	2006-2008
UF College of Liberal Arts and Sciences Anderson Scholar of High Distinction	Spring 2007
University of Florida University Scholar Recipient	2007 - 2008
International Baccalaureate Diploma	May 2005
Naval Science Award	2001
Alachua Region Science Fair – Grand Overall Champion	2001

ACTIVITIES

Martial Art Training – Shotokan Karate 1997 - 2003
A style of karate developed in Okinawa.
Promotes respect, compassion, humility, patience, and both an inward and outward calmness.

Ashtanga Yoga Fall 2007 – Present
Union of breath, body, and mind restores the system to its primal freshness.
Builds flexibility, strength and balance both physically and mentally.
Yoga chitta vritti nirodhah (Sanskrit): yoga is cessation of the fluctuations of the mind.

High School Mentoring Program 2007 -2010
Assisted Howard Bishop Middle School students with math and science skills.

Rubik's Cubes and Geometric Puzzles 2002 - Present
Hones pattern recognition, spatial memory, and improves finger dexterity.
Develops the ability to visualize and manipulate 3D objects in the mind's eye.

ATHLETIC AWARDS

National HS Swim Coaches Assoc. Swimming All-America,	2004 – 2005
Eastside High School National High School Heisman Candidate	2004 – 2005
Eastside High School Athlete Hall of Fame	2005
ABC Television 20 - Scholar Athlete of the Year	2004 – 2005
Two Time Gainesville Sun Swimmer of the Year	2003 – 2005
Eastside High School Swim Team Captain	2004 – 2005
Four Time United States Swimming Scholastic All – American	2001 – 2005
Four Time Finalist State High School Swimming Championships	2003 – 2005
Four Time Regional Champion FHSAA Finals Series	2003 –2005